

THESIS / THÈSE

MASTER DE SPÉCIALISATION EN INFORMATIQUE ET INNOVATION

Approche Open Data au Bureau Economique de la Province de Namur

Guilmin, Cédric

Award date:
2017

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



*Mémoire présenté en vue de
l'obtention du titre de*
**Master de spécialisation en
Informatique et Innovation**

Mémoire encadré par :
Professeur Naji Habra et
Bertrand Verlaine

ANNEE ACADEMIQUE 2016-2017

Remerciements

Ce mémoire n'aurait pas pu prendre forme sans les conseils avisés et la bienveillance de Mr. Habra. Je le remercie pour le temps qu'il m'a accordé. Je tiens également à remercier l'ensemble des professeurs du master BAGI ayant œuvrés à faire de nous des Business Analyst.

Deuxièmement, je remercie du fond du cœur les membres du Bureau Economique de la Province de Namur qui ont fait preuve d'une disponibilité et d'une flexibilité sans égales. Ils étaient à l'écoute de mes remarques et nous avons, grâce à cette expérience, construit une relation de confiance et de respect. Merci à François, Marc, Jean-Yves et Thibaut.

Je dis aussi un « Mille mercis » à mon entourage, à savoir Lindsay, mes parents et mes amis.

Table des matières

1	Introduction : contexte et description du BEP	1
1.1	Mandat.....	1
1.2	Description de l'organisation	1
1.2.1	Vue générale sur le BEP	1
1.2.1.1	Historique du BEP	1
1.2.1.2	Description actuelle du BEP	2
1.2.1.3	Missions actuelles	2
1.2.2	Culture organisationnelle du BEP	2
1.2.3	Organigramme	3
1.3	Motivation de la présence d'un Business Analyst	4
2	Définition de la stratégie du BEP sur l'Open Data	5
2.1	Introduction et description du processus.....	5
2.2	Stratégie du BEP.....	5
2.2.1	Missions génériques.....	5
2.2.2	Stratégie pour l'Open Data	6
2.2.2.1	SmartCity by BEP.....	6
2.2.2.2	Stratégie Open Data.....	7
2.2.2.2.1	Stratégie à court terme	7
2.2.2.2.1.1	Phase de communication/promotion au sein du BEP et vers les communes	7
2.2.2.2.1.2	Phase d'expérimentation au sein du BEP	10
2.2.2.2.2	Stratégie à moyen terme	13
2.2.2.2.2.1	Positionnement du BEP en matière d'Open Data vis-à-vis des communes	13
3	Open Data	15
3.1	Plateformes Open Data.....	15
3.1.1	Vision du BEP sur une plateforme Open Data	15
3.1.2	Analyse des besoins	15
3.1.2.1	Dynamique de la récolte des besoins	15
3.1.2.2	Identification des besoins	15
3.1.2.2.1	Besoins fonctionnels	15
3.1.2.2.1.1	Must have	16
3.1.2.2.1.2	Should have.....	20
3.1.2.2.1.3	Could have	20

3.1.2.2.2	Besoins non fonctionnels.....	21
3.1.3	Analyse des plateformes.....	22
3.1.3.1	Contexte et stratégie IT au sein du BEP.....	22
3.1.3.2	Approche utilisée pour l'analyse des plateformes.....	22
3.1.3.2.1	Analyse préliminaire sur CKAN	23
3.1.3.3	Socrata	24
3.1.3.3.1	Description générale.....	24
3.1.3.3.2	Besoins fonctionnels	24
3.1.3.3.2.1	Must have	24
3.1.3.3.2.2	Should have.....	27
3.1.3.3.2.3	Could have	27
3.1.3.3.3	Tarification	27
3.1.3.4	OpenDataSoft.....	28
3.1.3.4.1	Description générale.....	28
3.1.3.4.2	Besoins fonctionnels	28
3.1.3.4.2.1	Must have	28
3.1.3.4.2.2	Should have.....	31
3.1.3.4.2.3	Could have	32
3.1.3.4.3	Tarification	32
3.1.4	Comparaison des plateformes et recommandations	33
3.1.4.1	Discussion des véritables différences entre les plateformes.....	33
3.1.4.2	Recommandation.....	34
3.2	Prochaines étapes.....	35
3.2.1	Licences	35
3.2.2	Cadres réglementaires	36
3.2.2.1	Réutilisation des données du secteur public	36
3.2.2.2	Protection des données à caractère personnel	36
4	Conclusion.....	39
5	Bibliographie	40

1 Introduction : contexte et description du BEP

1.1 Mandat

Dans le cadre de la réalisation de mon mémoire, j'ai contacté le Bureau Economique de la Province de Namur début décembre 2016 afin de voir s'il était possible que je puisse apporter mon aide sur un projet. J'ai rencontré dans un premier temps François Laureys (gestionnaire de projet en stratégie territoriale et Smart Cities). Il m'a présenté le BEP ainsi que les différents projets sur lesquels il travaille. Il m'a présenté la stratégie du BEP en matière de Smart Cities et ses premières réflexions concernant l'Open Data. Je sentais que c'était sur ce dernier domaine que je pouvais apporter ma pierre à l'édifice car je suis intéressé par le domaine des données. Mais l'Open Data en tant que tel m'était jusqu'à lors inconnu.

Lors de ma deuxième réunion avec le BEP, j'ai rencontré Marc Deroitte (directeur du département informatique) et Jean-Yves Pagès (membre du département informatique). Beaucoup de points furent soulevés : le scope du projet n'était pas encore défini. Après un temps de réflexion et en accord avec les membres du BEP, nous avons décidé d'inclure dans le scope du projet les 2 éléments suivants :

- Définition d'une stratégie potentielle et des éléments de réflexions quant au positionnement du BEP dans le paysage de l'Open Data ;
- Evaluation de plateformes Open Data et recommandation d'une plateforme qui intégrerait les objectifs et besoins poursuivis.

L'idée de travailler dans la ville où j'ai toujours vécu me semblait une opportunité d'avoir une vue sur les projets en cours et apporter ma part de contribution à son développement. De plus, j'ai ressenti, dans le chef des parties prenantes, une bienveillance, une volonté d'avancer ainsi qu'une envie de partager leur expérience. Ce contexte me semblait donc favorable à une collaboration et c'est pourquoi j'ai décidé de me lancer dans l'aventure avec le BEP.

1.2 Description de l'organisation

1.2.1 Vue générale sur le BEP

1.2.1.1 Historique du BEP

Le Bureau Economique de la Province de Namur fut créé le 14 décembre 1962, sous la forme d'une ASBL. Le but du BEP est, au début de son existence, de redynamiser le développement industriel : il l'effectuera en favorisant la création de zonings sur lesquels, entre autres, de grandes entreprises étrangères viennent s'installer [1]. Le BEP aidera également la province avec la construction de logement, l'aéroport de Namur, le domaine de Chevetogne, etc [1].

Au 1^{er} janvier 1977, la fusion des communes en Belgique fait passer leur nombre d'environ 2400 à un peu moins de 600 : la taille de Namur ainsi que son nombre d'habitants augmentent automatiquement [1]. En 1979, le BEP devient une intercommunale. C'est un peu moins de 10 ans après, en décembre 1986, que Namur est décrétée capitale de la Région Wallonne [2]. Namur et la province de Namur prennent donc une importance plus grande dans le paysage politique et économique. L'objectif du BEP n'est plus d'aider uniquement les grands groupes étrangers, mais d'aider les entreprises implantées chez nous et de tisser le réseau de PME [1]. Les communes se situant les plus au sud de la province sont revalorisées.

Vers la fin des années 1980, début des années 1990, le BEP noue des partenariats avec le monde scientifique [1] : la Faculté agronomique de Gembloux ainsi que les Facultés universitaires Notre-Dame de la Paix. De 1993 à 2003, le BEP se modernise pour faire face à un monde fait de concurrence. Ce sont les différents départements propres à chaque métier qui apparaissent : développement économique, développement territorial et gestion de l'environnement [1].

1.2.1.2 Description actuelle du BEP

Le BEP est une intercommunale : son existence résulte de l'association de 38 Communes de la province de Namur [3]. Le but d'une intercommunale est de créer une structure commune pour un ensemble de matières générique à l'ensemble des communes : la gestion des déchets, l'animation et la gestion du territoire, ... La centralisation de certaines compétences permet des économies d'échelle ainsi que de partager une expérience commune.

Le BEP a, comme une grande majorité des intercommunales, opté pour le statut de S.C.R.L (Société Coopérative à Responsabilité Limitée). Son actionnariat est exclusivement composé des différentes communes de la Province de Namur.

1.2.1.3 Missions actuelles

Le BEP est en charge des missions suivantes [3] :

- Développement économique
- Développement territorial
- Environnement
- Palais des Expositions
- Crematorium
- Energie

Un ensemble de départements viennent supporter la structure opérationnelle : Ressources Humaines et Services Généraux, Comptabilité et Finances, Informatique et Télécommunications, Relations avec les administrateurs et les associés, Communication. [3]

1.2.2 Culture organisationnelle du BEP

Le BEP est une intercommunale qui se veut moderne. En effet, elle a été une des premières à proposer le télé-travail à ses collaborateurs. Elle accorde une grande importance quant à l'autonomie laissée à ses collaborateurs : le temps ainsi que le lieu de travail sont des paramètres flexibles que le collaborateur peut adapter en fonction de ses besoins. Le BEP offre d'ailleurs un environnement de travail très chaleureux et propice à la création (notamment la pièce « Créaroom »). L'identité du BEP se propage sur tout l'organigramme et il obtient en retour un engagement fort de ses travailleurs. Cela est possible grâce à un management qui se veut innovant et proches des actions menées par ses équipes.

Le BEP prône un ensemble de valeurs qui sont les suivantes [3] :

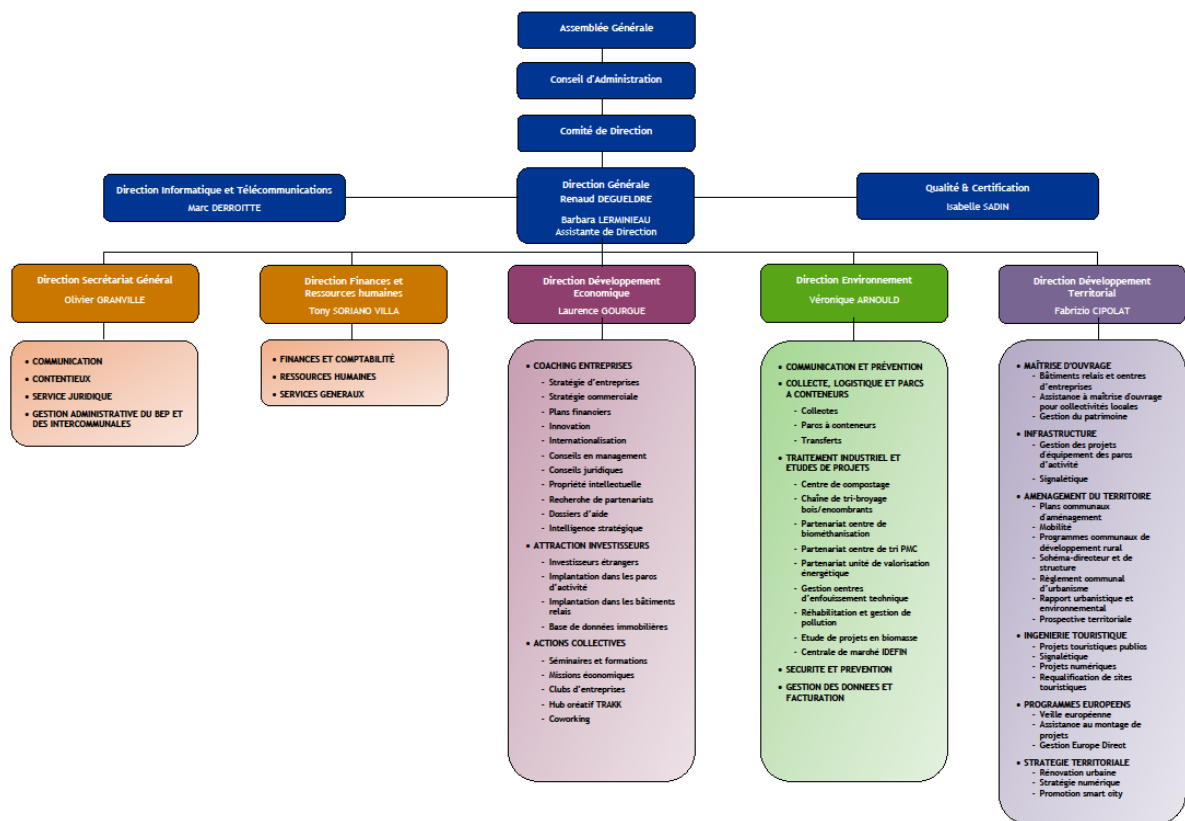
- Transparence : le BEP, soucieux d'une bonne gouvernance, montre l'exemple en se tenant à la disposition des mandataires et des citoyens pour montrer et expliquer ce qu'il entreprend et les moyens qui y sont accordés. Cette transparence réapparaît lorsque la stratégie Open Data sera esquissée dans la section « 2.2.2 Stratégie pour l'Open Data ».
- Orientation client (les clients sont les communes) : le BEP est un organe pluridisciplinaire qui, grâce à son expertise en interne et son réseau de professionnels, peut rendre une multitude de

services aux communes. Il se veut compétent et rigoureux car la satisfaction des clients est un élément essentiel.

- Autonomie et responsabilisation des collaborateurs : le BEP est empreint d'un management innovant et il favorise la collaboration ainsi que la co-crédation en ses murs.
- Recherche de solutions et remise en question de ses « acquis » : le BEP a compris que la société se digitalise et se numérise de plus en plus et elle a décidé d'intégrer au fur et à mesure cet aspect dans l'offre de services. Être proactif fait partie de la démarche d'innovation.

1.2.3 Organigramme

L'organigramme du Bureau Économique de la Province de Namur est présent ci-dessous. Les départements « business » sont le « Développement Économique », l'« Environnement » et le « Département Territorial ». Les départements de support sont les suivants : « Informatique et Télécommunications », « Secrétariat Général » (Communication, Contentieux, Services Juridiques et Gestion administrative du BEP et des intercommunales) et « Finances et Ressources Humaines » (Finances, Ressources Humaines et Services Généraux).



1.3 Motivation de la présence d'un Business Analyst

Ma présence est motivée par plusieurs aspects. Lors des premières rencontres avec les personnes impliquées dans ce projet, j'ai remarqué qu'il y avait peut-être un besoin en Open Data mais ce besoin était flou et personne ne savait réellement par où commencer, à quoi faire attention. L'incertitude était grande quant à la nécessité et la direction prise par le projet. J'étais également dans la même situation et, en tant que Business Analyst, j'ai passé une grande partie de mon temps à faire progresser mon « domain understanding » : je me suis beaucoup documenté sur le sujet afin de pouvoir guider le projet. J'ai travaillé de manière méthodique et structurée pour pouvoir rassurer les parties prenantes et me rassurer par la même occasion : toute la difficulté fut de travailler dans ce cadre incertain, avec ce « problème » non-structuré, pour arriver vers une situation davantage définie et organisée.

En proposant des solutions ou des pistes de réflexions, on a pu voir que l'adage IKIWISI (« I know it when I see it ») prenait tout son sens. C'est en montrant et proposant des choses que les gens ont été plus à même de définir ce qu'il voulait ou ce qu'il ne voulait pas. J'ai travaillé en étroite collaboration avec les personnes du BEP et nous étions dans un processus d'itérations : plus on avançait, et plus on était capable de déterminer ce dont nous avions besoin.

Dans le contenu à proprement parler du projet, j'ai dû faire attention à garder le cap entre une stratégie « high-level » et son implication au niveau opérationnel. Le business analyst doit être capable de travailler à différents niveaux et de pouvoir communiquer avec des personnes du management, des personnes du business et des personnes techniques. Au cours de ce projet, pour pouvoir communiquer avec les personnes du management et les personnes du business, je me suis beaucoup informé sur le domaine de l'Open Data ainsi que la manière dont le BEP travaille. Concernant l'aspect plus techniques et l'aspect données, j'ai compté sur mon expérience en Business Intelligence, en analyse de données et en data warehousing : j'ai utilisé, entre autres, ces connaissances pour extraire le maximum d'un projet Open Data et pour mettre en place un cadre permettant la réutilisation la plus efficace de ces données.

2 Définition de la stratégie du BEP sur l'Open Data

2.1 Introduction et description du processus

L'objectif de la section « 2.2 Stratégie du BEP » est de décrire de manière succincte la stratégie générale pour le BEP, la section « 2.2.2.1 SmartCity by BEP » décrit la stratégie du BEP dans le cadre des Smart Cities et la section « 2.2.2.2 Stratégie Open Data » explique en quoi pourrait consister la stratégie en matière d'Open Data et comment elle s'inscrirait dans la stratégie « SmartCity by BEP » et dans la stratégie générale du BEP.

Premièrement, dans l'optique de se convaincre de la valeur ajoutée de l'Open Data, il faut pouvoir construire un argumentaire, un socle de convictions sur lesquels le BEP doit forger son identité avant d'évaluer les stratégies à adopter par le BEP autour de la thématique Open Data. Cela permettra clairement de définir le but, le sens de toutes nos actions entreprises dans le cadre de ce projet. Ce point consistera à approcher les communes pour expliquer ce qu'est l'Open Data, en quoi (et pourquoi) il peut nous aider et comment « on fait de l'Open Data ».

Dans la stratégie Open Data, il faut pouvoir également positionner le BEP par rapport aux services potentiels qu'il pourrait rendre. Cette proposition de positionnement permettra de définir le rôle que le BEP prendrait.

2.2 Stratégie du BEP

2.2.1 Missions génériques

La mission principale du BEP est de créer une dynamique économique et sociale durable [3]. Les missions génériques pour l'ensemble du BEP pourraient être résumées comme suit [4] :

- Développement économique et entrepreneurial : le BEP a pour mission de soutenir le développement économique pour les différents acteurs (entreprises, organismes, associations et citoyens.) ;
- Amélioration du cadre de vie : il est important de rendre le territoire namurois plus agréable à vivre en innovant et en respectant ses spécificités.

En ce qui concerne la manière de réaliser ses missions, le BEP a pour objectif d'être « un modèle d'entreprise publique » [4]. Dans ce cadre, les éléments suivants, en cohérence avec sa culture organisationnelle, soutiennent le « comment » dans sa stratégie :

- Gouvernance empreint de transparence vis-à-vis des parties prenantes;
- Autonomie et responsabilisation de ses travailleurs grâce à un management innovateur ;
- Recherche de solutions avec proactivité, anticipation, souplesse et réactivité de ses équipes.

Ces dernières années, des scandales ont entachés plusieurs intercommunales dont le mode de fonctionnement était inconnu du grand public, peu transparent dans l'objectif d'en faire bénéficier certains. Le BEP veut être irréprochable et le management a fait de la transparence un objectif important. D'ailleurs, le management du BEP a convié ses collaborateurs à une réunion en début d'année 2017 pour les conscientiser sur enjeux liés à la transparence.

Un exemple en matière de transparence est la commune d'Issy-les-Moulineaux, située dans le département des Hauts-de-Seine en France. Issy-les-Moulineaux a commencé sa plateforme Open Data début 2015. La ville a développé une offre de services numériques et cela est représentatif du terreau

dans lequel les entreprises peuvent grandir : la ville accueille le siège de Microsoft France, le siège de Cisco System ainsi qu'une multitude d'entreprises présentes dans le domaine des technologies de l'information et de la communication. La ville d'Issy-les-Moulineaux s'est focalisée sur l'intelligibilité de ses données sur le fond et sur la forme qu'elles prennent : ils partent du principe que tout citoyen ne doit pas être un technicien pour comprendre et exploiter ces données. Ce premier aspect a donc entraîné une émergence de la réutilisation de l'Open Data. Issy-les-Moulineaux a donc réussi un double objectif. Le service des finances de la ville d'Issy-les-Moulineaux a mis à disposition les données financières pour favoriser sa transparence ainsi que l'intelligibilité à travers une plateforme où la facilité d'utilisation ainsi que le 'look and feel' sont les maîtres mots.

Dans le plan stratégique du BEP pour 2017 [4], le Développement Territorial ainsi que le Développement Économique sont marqués par le numérique. Le département 'Développement Économique' a pour mission d'intégrer les entreprises dans le secteur du numérique. La digitalisation a pour effet de d'accélérer la croissance d'une entreprise via divers moyens : personnalisation de l'offre, public cible plus large, augmentation de la productivité, diminution des coûts, ... Les business models qui y sont relatifs doivent aussi être repensés, refondus pour préciser ce qui fera réellement la valeur ajoutée de l'entreprise et, par voie de conséquence, ce qu'elle pourra vendre.

Le Développement Économique a également pour mission de favoriser la création de produits en lien avec les Smart Cities : c'est donc ici que ce département rencontre le Développement Territorial. Ce dernier est en charge de la démographie, le climat et l'énergie, le territoire ainsi que des différentes utilisations du territoire. Une démarche de Smart City permet de trouver des solutions innovantes en tenant compte de ces missions, de ces enjeux.

2.2.2 Stratégie pour l'Open Data

2.2.2.1 SmartCity by BEP

Le Bureau Economique de la Province (BEP) de Namur se positionne sur la thématique de la SmartCity au travers de son programme d'actions 'SmartCity by BEP'. Dans ce cadre, ils ont développé différents aspects qui peuvent être abordés lorsqu'on parle de 'Smart City' ('Smart Education', 'Smart Healthcare', 'Smart Mobility', 'Smart Energy', ...). Ces différentes thématiques peuvent être soutenues et accélérées grâce au partage de données (Open Data).

Le slogan 'Le territoire Namurois plus ingénieux' révèle en fait plusieurs aspects sur la volonté du BEP de rendre le territoire namurois [5] :

- Plus ingénieux
- Plus collaboratif
- Plus impliquant
- Plus attractif

Les différentes actions menées par le BEP [5] sont :

- Animation du territoire : le BEP mène des actions d'information à l'intention des communes pour sensibiliser ce que les technologies de l'information et de la communication peuvent apporter un terme de solutions à des problématiques d'environnement, de mobilité, ...
- Compréhension et valorisation du territoire (expérimentation et mise à disposition) : le BEP a pour mission de comprendre les spécificités de chaque territoire. Le BEP expérimente et essaie de

trouver des solutions. « Le BEP vise à se positionner comme pilote de projets innovants ayant une vocation supra-communale » [4]. Un exemple a été de mettre en place une mobilité plus accessible pour les zones rurales en utilisant technologie et collaboration entre citoyens.

- **Dynamisation du territoire** : le BEP est consulté lorsqu'une commune veut entreprendre un projet. Le BEP intervient en tant qu'assistant maître d'ouvrage et il assure le bon suivi du projet entre ce que demande les communes et le maître d'œuvre. Dans ce contexte, le BEP propose une série de services aux communes.

2.2.2.2 *Stratégie Open Data*

2.2.2.2.1 *Stratégie à court terme*

2.2.2.2.1.1 *Phase de communication/promotion au sein du BEP et vers les communes*

« A smart city uses information and communication technologies (ICT) to enhance quality, performance and interactivity of urban services, to reduce costs and resource consumption and to improve contact between citizens and government. [...] The key to making a city smart is to centre the initiative on people and openness. The citizens create the culture that shapes the city's characteristics, including the sense of community and the diversity; thereby ensuring the collective wellbeing of its citizens. Open Data is a key feature in most smart city initiatives. » [6]

On peut donc voir l'Open Data comme un enabler, un facilitateur d'une stratégie Smart City : établir des projets Smart City sans avoir de l'Open Data réduit l'impact de ces derniers. En expérimentant, le BEP développe ses compétences et il acquiert une expertise.

En montrant l'exemple, le BEP veut montrer le :

- **QUOI**. Démystifier ce qu'est l'Open Data. De façon générale, le monde dans lequel on vit est de plus en plus connecté, de plus en plus générateur de données. Les administrations doivent pouvoir comprendre cette nouvelle dynamique et voir qu'après tout, l'Open Data n'est pas une boule de cristal, ni une discipline obscure. Lorsqu'on explique ce qu'est l'Open Data, on s'intéresse au « quoi ». Avant même d'être un mot à la mode lorsqu'on évoque la thématique de 'Smart Cities', l'Open Data est d'abord un mouvement qui vise à partager les données : tout le monde peut y accéder, les utiliser et les partager [7]. Par ouverture des données, on entend 3 concepts phares :
 - **Accessibilité** : l'utilisateur a accès à la donnée. En d'autres mots, il existe une plateforme très facilement accessible pour les citoyens.
 - **Utilisabilité** : le format de stockage des données s'adresse à tous les citoyens, quel que soit leur niveau technique. La donnée est, aussi bien dans le fond que dans la forme, facilement utilisable : cela permet à l'utilisateur de comprendre les données qu'ils traitent (cela est possible si des métadonnées fiables sont également à la disposition de l'utilisateur).
 - **Gratuité** : l'utilisation, la modification et la redistribution des données par n'importe quel individu ou entité sont gratuites [8]. Ces sets de données seront accompagnés de licences en continuité avec cette philosophie d'ouverture et cela lèvera toute ambiguïté et confusion quant à ce que les utilisateurs pourront faire avec les données.
- **POURQUOI**. Le BEP explique les retombées économiques, politiques et sociales lors de l'ouverture et du partage de données publiques. Le BEP veut convaincre de faire de l'Open Data une

thématique à part entière au sein d'une administration. L'impact de l'Open Data doit également être monitoré pour mesurer son ampleur par rapport aux données mises à disposition : le rapport 'Open Data Maturity' publié par l'European Data Portal [8] fait état qu'il y a une amélioration quant à la mise en place d'une approche systématique de mesure de l'impact mais certains pays se situent encore en arrière sur cet aspect : le BEP doit faire attention de ne pas passer à côté de ce point. En couplant les initiatives Open Data à une approche pour mesurer ces impacts, le BEP sera en mesure d'évaluer précisément et objectivement ce qui est réalisé. Ces facteurs mesureront/évalueront :

- le succès et l'engouement d'une politique Open Data [8] ;
 - les facteurs de résistance et/ou de non-intérêts grâce aux feedbacks des utilisateurs ;
 - les prérequis nécessaires à la réussite d'une stratégie d'ouverture des données.
- **COMMENT.** Le but est de montrer la direction à suivre aux administrations désireuses de partager leurs données. Via cet aspect, on s'intéresse davantage au « comment ». On explique les moyens à dispositions pour aider au partage et à la diffusion de ces données. Dans la stratégie pour 2017 [4], l'accent est mis sur l'ouverture des données propres au BEP. Ce dernier peut donc montrer l'exemple et cela aura pour but d'atteindre un double objectif : convaincre les communes et montrer la transparence du BEP via cette démarche. En optant pour l'ouverture des données via une plateforme Open Data, les données seront disponibles au grand public avec des outils modernes et à la portée de tous. Si des études plus avancées sont nécessaires pour répondre à une problématique déterminée, le « comment » sensibilise aussi au fait que des compétences spécifiques sont nécessaires : ces compétences permettent d'extraire encore plus de valeur des données (data mining, machine learning, predictive analytics).

Dans la suite de cette section, nous allons davantage nous intéresser au **POURQUOI** et expliquer les bénéfices à proposer des données ouvertes.

L'estimation de l'impact de l'Open Data sur la société est très difficile à analyser et à quantifier. Cela est dû à plusieurs facteurs [8] :

- identifier l'effet causal, c'est-à-dire identifier la cause et la relier avec les conséquences sur la vie sociale, économique ou politique ;
- l'effet de l'ouverture des données ne s'arrête pas à un moment déterminé dans le temps : l'effet est distillé dans le temps ce qui rend la mesure souvent sous-estimée par rapport à l'effet réel ;
- les effets ne sont pas uniquement monétaires : l'ouverture des données peut servir à soutenir une politique spécifique menée par un gouvernement, une autorité locale (suppression des inégalités entre les sexes, information sur les infrastructures en direction des personnes à mobilité réduite, inclusion du citoyen dans les discussions concernant les finances d'une administration, informer le citoyen pour qu'il améliore sa qualité de vie, ...). L'Open Data alimente l'économie de la connaissance également [9].

Les impacts peuvent être de plusieurs types : il y a des impacts sociaux, économiques et politiques.

Au niveau des citoyens

Le partage et l'étude de ces données permettent une meilleure connaissance du territoire : les décisions des administrations sont davantage en phase avec ce que les citoyens vivent ou ressentent. En Espagne par exemple [8], Sin Barreras, un moteur de recherche, recense une série d'endroits ainsi que leur niveau

d'accessibilité : les personnes à mobilité réduite peuvent dès lors s'informer quant aux endroits qui leur semblent accessible compte tenu de leurs difficultés.

La qualité de vie des citoyens est accrue grâce à des villes plus sûres, moins congestionnées et plus propres. La ville de Paris a mis sur pied une démarche visant à analyser des points très congestionnés de la ville : Paris veut « augmenter l'attractivité de la ville et réduire son empreinte carbone » [10]. La ville de Paris a placé toute une série de capteurs afin de mieux comprendre les flux régissant des places où la circulation est dense (Place de la Nation, ...). Ces données sont disponibles en temps réel sur la plateforme Open Data de la ville de Paris ainsi que sur des écrans tactiles. En comprenant comment la dynamique d'une place fonctionne, la ville de Paris sera en mesure d'améliorer l'expérience des gens utilisant cette place (diminution du trafic, diminution de la pollution atmosphérique et sonores, diminution des accidents).

Par exemple, une approche Open Data peut servir à mesurer la consommation des bâtiments. L'Open Data pourrait être utilisée dans une application concrète présente au sein de la stratégie 2017 du BEP [4] : un projet de comptabilité énergétique automatisée est à l'étude et ces données pourraient être collectées et mises à disposition pour mesurer et monitorer la consommation de certains bâtiments.

Au niveau politique

Les pouvoirs publics sont tenus à une gouvernance interne plus importante grâce à une transparence favorisée par l'Open Data. L'ouverture des données permet d'avoir un regard sur les processus et les événements qui se passent au sein des administrations : l'Open Data est également un facilitateur pour combattre la corruption [8]. Cette transparence, couplée avec des initiatives où les citoyens sont amenés à collaborer – projet en cours dans le 'SmartCity by BEP' –, fait augmenter la participation et l'engagement des habitants dans des projets citoyens. Cette transparence n'est atteinte que lorsque les citoyens ont les moyens de comprendre et/ou d'utiliser ces données [11], d'où à nouveau le travail nécessaire de conscientiser et de montrer ce qui peut être accompli avec les données.

Il y a un gain en efficacité des administrations : centraliser les données permet d'avoir un point névralgique où gérer la qualité de données. Tout le monde peut bénéficier de cette centralisation en utilisant ces données et ces réutilisateurs en bénéficient plus rapidement. De plus, les actions menées par les administrations sont menées aux endroits où cela est le plus nécessaire, le plus urgent [11].

Le rapport 'Open Data Maturity in Europe 2016' [8] mesure le degré de maturité des politiques menées et il évalue les fonctionnalités présentes sur les 28 portails nationaux. Ce rapport fait également une étude qualitative quant au ressenti de l'impact par les autorités : il fait état, entre autres, que les données liées à la mobilité (circulation, transports en commun) ont un grand impact sur la vie des citoyens, sur l'organisation de la ville ainsi que sur les pouvoirs politiques. Les données liées au transport font parties des données les plus demandées [6].

Au niveau économique

L'estimation économique de l'Open data est très difficile à cause aussi des effets indirects que provoque l'Open Data. Il faudrait mesurer quelle est la part de la valeur ajoutée d'un service apportée grâce à l'utilisation de l'Open Data. Cette part vient de l'augmentation de la valeur ajoutée et/ou de la diminution des coûts (ou du gaspillage) liés à la production d'un bien, d'un service. De nouveaux services ainsi que des nouveaux business models – de l'innovation de manière générale - émergent grâce à l'Open Data.

Le rapport 'Creating Value through Open Data' indique que le marché Open Data va croître, en Europe, de 36,9% entre 2016 et 2020 créant 25 000 jobs dans ce secteur et 75 000 jobs grâce à la réutilisation des données [9] : les PME sont des entreprises où le data-driven innovation est plus souvent utilisé, ce sont donc celles-ci qui sont le plus à même de générer des emplois. Selon les estimations, la valeur des effets directs représente seulement 1/3 de la valeur des effets indirects.

Les études montrent que le return on investment (ROI) observé dans des projets Open Data est toujours positif par rapport à l'investissement initialement consenti [9] : le ROI pour l'Union Européenne des 15 pays en année 2000 avait été estimé à 7 tandis que celui des États-Unis d'Amérique a été estimé à 39. En d'autres mots, 1€ investi dans l'Open Data aux États-Unis créait 39€ pour l'économie. Le rapport 'Creating Value through Open Data' explique que des études ont été menées quant à l'effet de la tarification des données ouvertes sur la croissance d'un secteur et il en ressort que plus la tarification se rapproche de 0, plus la réutilisation et l'impact des données sur l'économie sont grands [9].

2.2.2.2.1.2 Phase d'expérimentation au sein du BEP

Le BEP Namur veut **montrer l'exemple** [5] : il a lancé son site d'Open Data durant le mois de décembre 2016. Le BEP innove et expérimente dans sa démarche : il est un pionnier au niveau de la Province de Namur quant à la réflexion d'une stratégie d'Open Data ainsi que des solutions concrètes en ligne avec ces objectifs. En expérimentant, le BEP permet d'évaluer les compétences qui seraient nécessaires tout en se construisant une expérience : durant cette phase d'approche, le BEP évaluera au cas par cas ce qu'il peut réaliser et pourquoi il pourrait le réaliser. Cette réalisation peut s'effectuer en tout ou en partie, les parties non réalisées pouvant être externalisée par exemple.

Au fur et à mesure de la mise en place de la stratégie court terme, le BEP doit analyser plus profondément le rôle exact qu'il veut jouer dans l'Open Data et son positionnement par rapport aux communes ainsi qu'à la région. C'est durant cette phase que des éléments plus concrets seront organisés pour construire et intégrer une offre et un positionnement avec les parties prenantes actives sur la digitalisation et l'amélioration de la qualité de vie dans la Province de Namur. Les éléments de réflexions suivantes – non exhaustifs – seront à analyser et à évaluer si le BEP se placerait dans l'hypothèse d'une offre de services Open Data aux communes :

- Le BEP serait une option intéressante pour offrir des services Open Data pour toutes les communes en leur permettant un coup d'entrée en la matière plus bas : grâce à des économies d'échelle, les coûts fixes pourraient être répartis entre les différentes communes. Il est encore à évaluer ce qui est réalisé par le BEP de ce qui pourrait être externalisé.
- Le BEP serait un partenaire privilégié vu les communes ont déjà transféré la gestion courante de certaines de leurs missions : le BEP a la confiance des communes et il a une structure qui se professionnalise en fonction des services à rendre.
- Il ne serait pas illogique d'attribuer la mission de l'Open Data au BEP qui pourrait, par la même occasion, faire bénéficier les communes des compétences que le BEP dispose en son sein. Ces compétences dépendraient des données à mettre à disposition (finance, économie, aménagement du territoire, ...).
- Si le BEP était le point central de l'Open Data pour la province de Namur, il pourrait améliorer sa connaissance du territoire en proposant une ligne intégratrice sur les données mises à disposition. Le BEP pourrait donc étudier la province de Namur comme une seule entité ainsi que comparer

les communes : ces agrégations et ces comparaisons sont rendues possibles seulement si une méthode de travail commune est appliquée lors de la collecte d'information auprès des communes. Le concept sous-jacent serait une interopérabilité de plusieurs jeux de données [12].

Dans l'ère de la digitalisation et de l'ouverture de données, le BEP – entreprise publique et moderne – aurait l'ambition, à terme, de s'inscrire dans le mouvement de l' « Open Government Data » qui a pour objectif de rendre les administrations plus transparentes et plus accessibles. Dans l'optique d'ouverture des données du BEP, les concepts généraux de l' « Open Government Data » sont en fait très concrets et sont un support important pour comprendre et réussir une démarche Open Data [13] (*les concepts sont laissés en anglais pour ne pas dénaturer leur signification*):

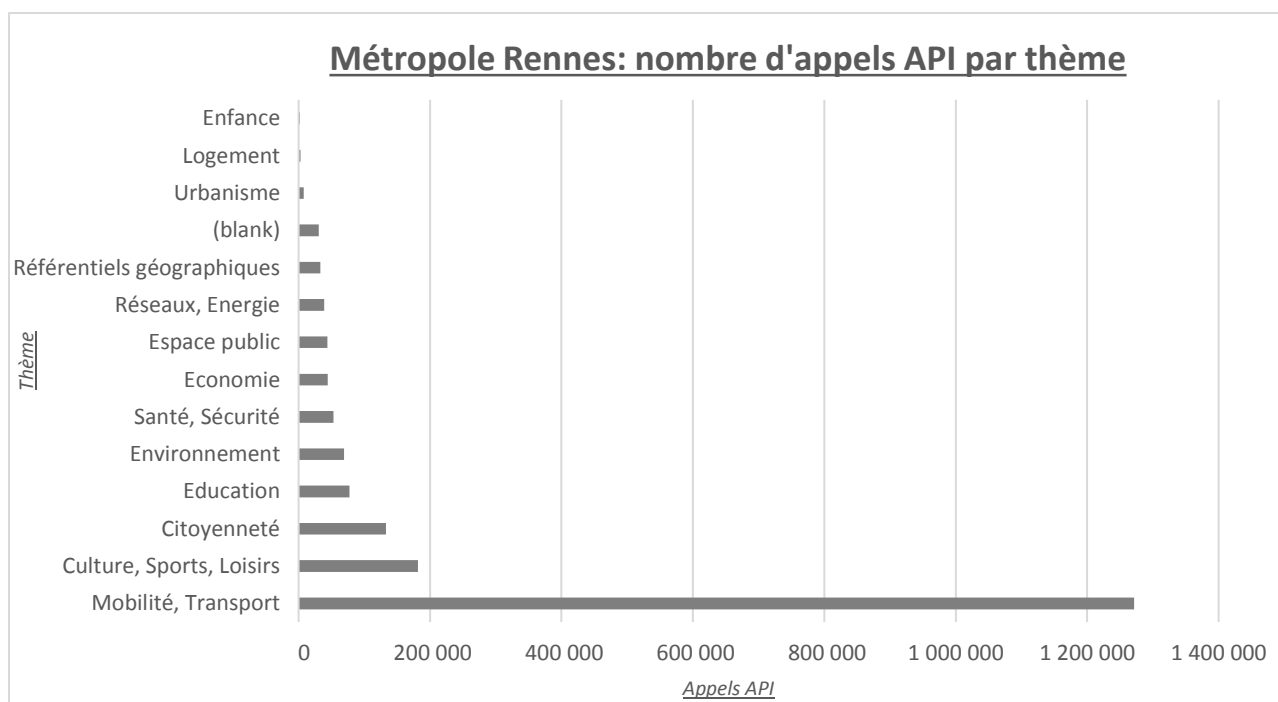
- **Complete** : toutes les données publiques – sauf celles qui violeraient la vie privée ou la sécurité – sont mises à disposition.
- **Primary** : les données sont le plus brut dans le sens où elles sont présentées avec le plus fin niveau possible de granularité (pas d'agrégation).
- **Timely** : les données sont mises aussitôt que possible à disposition.
- **Accessible** : le plus grand nombre d'utilisateurs a accès aux données à des fins aussi diverses que possible.
- **Machine processable** : les données sont suffisamment structurées que pour être traitées par une machine.
- **Non-discriminatory** : les données sont à disposition de tout le monde, sans inscription nécessaire.
- **Non-proprietary** : le format des données n'est pas sous le contrôle d'une entité particulière.
- **License-free** : il n'y a pas de copyright, de brevets, de marques déposées ou de secrets commerciaux. Néanmoins, les données sont soumises à des licences ouvertes favorisant la réutilisation des données : il y a donc peu d'obligations que le réutilisateur de données doit satisfaire (ce point sera abordé plus tard dans le travail).

European Data Portal offre une grille de lecture intéressante lorsqu'il évalue la maturité des stratégies Open Data des différents pays [8] [6]. Le BEP peut s'inspirer de ces constations pour voir s'il doit y avoir des pistes à privilégier parmi les éléments avancés par l'European Data Portal lorsqu'il évaluera la possibilité de mettre ses propres données à disposition tout comme la possibilité d'offrir des services Open Data aux communes. Les points suivants sont là pour lancer la réflexion et la construction d'une stratégie qui formeront la ligne conductrice des actions ainsi que l'approche menée par les porteurs du projet.

- Utilisabilité du portail : portail « user-friendly » qui n'est pas uniquement destiné aux techniciens de l'information, possibilités de commenter les datasets, de demander des explications, ...
- Réutilisation des données : les données sont dans des formats facilement utilisables pour l'humain ainsi que pour la machine, les données sont de bonne qualité, ...
- Approche concernant les données à mettre à disposition : plusieurs approches sont possibles dans la manière d'évaluer ce qui va être mis à disposition. Chacune des approches proposées ci-dessous ne sont pas exclusives les unes des autres mais certaines sont poussées par des forces contraires.

Il est donc important de connaître quelle est la ligne directrice proposée par le management en la matière :

- Mise à disposition d'un maximum de jeux de données : la Grèce a opté pour cette stratégie durant l'année 2015-2016 [8].
 - Mise à disposition des données de haute qualité.
 - Mise à disposition des données à très haut impact : l'impact doit encore être évalué par rapport à quelle partie prenante considérée.
 - Mise à disposition de données demandée par les utilisateurs : cette approche est, entre autre, utilisée par des villes comme Barcelone, Copenhague et Vienne [6].
 - Mise à disposition de données pour répondre à des stratégies particulières : volonté de transparence des administrations, problématique spécifique d'une ville, amélioration des processus d'une administration, besoin de centralisation de données, création d'opportunités économiques, ...
- Domaines qui sont couverts : il faut peut-être regarder ce qui est le plus publié ainsi que le plus consulté, réutilisé. En ce qui concerne les domaines contenant le plus de datasets, le rapport 'Open Data Maturity' [8] mentionne que les 5 domaines les plus représentés sont par ordre décroissant d'importance 'Justice, systèmes légaux et la sécurité publique' (31%), 'Environnement' (21%), 'Sciences et technologies' (13%), 'Régions et villes' (11%) et 'Population et société' (6%). Ces 5 domaines représentent 82% de l'ensemble des datasets. Concernant les données les plus souvent consultées, il semblerait que les domaines soient très dissimilaires d'un pays à l'autre : cet aspect est à considérer durant la phase d'expérimentation menée par le BEP. Ci-dessous, un graphique représente le nombre d'appels à une API par thème pour la plateforme Open Data de Rennes Métropole. Par ce graphique, la répartition de l'utilisation des données, concernant un territoire donné, est également très disproportionnée en fonction du thème des données. Rennes ne fait pas figure d'exception lorsqu'on remarque que les données les plus utilisées concernent la thématique liée à la mobilité et du transport.



2.2.2.2.2 Stratégie à moyen terme

2.2.2.2.2.1 Positionnement du BEP en matière d'Open Data vis-à-vis des communes

Dans la démarche Open Data, le Bureau Économique de la Province de Namur pourrait se placer comme **un conseiller et un catalyseur**, un accélérateur du processus d'ouverture des données des administrations (communales). Durant la stratégie à court terme, le BEP va engranger l'expérience ainsi que le réseau spécifique au domaine de l'Open Data. Si une expertise est nécessaire pour des projets spécifiques, le BEP fera appel à son réseau pour maintenir la satisfaction des clients à un haut niveau.

Pour que le BEP devienne un acteur de référence dans la thématique Open Data, après avoir convaincu les communes de prendre le pas en cette direction, le BEP prouve qu'il est mesure d'apporter son expérience ainsi que son agilité dans l'accompagnement des communes. En effet, les choses évoluent très vite dans le domaine du digital et ce ne serait pas le moment :

- d'accumuler du retard au niveau de l'ouverture des données en Province de Namur ;
- de ne pas montrer un alignement dans la stratégie d'Ouverture des données avec celle définie au niveau fédéral et régional ;
- de multiplier les coûts et de manquer d'une approche intégrée lorsque des communes lancent leur propre initiative.

Le BEP pourrait accompagner les communes désireuses de partager leurs données. Cet accompagnement se ferait à 2 niveau :

- Au niveau du conseil (**conseiller**) : en guidant les autorités sur la manière de procéder lorsqu'une volonté d'ouverture des données est initiée. Ce conseil se composerait de divers services tels que
 - o le recensement des données,
 - o la phase de pré-étude quant à la forme et au contenu des données (qualité),
 - o un avis sur l'anonymisation des données,
 - o une évaluation de la fréquence de rafraichissement des données,
 - o une évaluation quant à la nécessité d'automatiser le processus de création et de mises à disposition des données,
 - o une stratégie pour évaluer l'impact des futures données à disposition.
- Au niveau opérationnel (**catalyseur**) : en mettant à disposition une plateforme Open Data offrant l'hébergement, la visibilité et la réutilisation facile des données.
 - o LE BEP offrirait un ensemble de service quant à la réalisation concrète et opérationnelle de la mise à disposition des données sur une plateforme. Le BEP prendrait en charge, voire formerait, les communes voulant partager leurs données. Le BEP s'assurait d'avoir une plateforme sur laquelle elle a la maîtrise et il assurerait une continuité de services en la matière.
 - o Le BEP, dans son portfolio de services, respecterait les guidelines et la gouvernance IT définie par le département IT.

Le BEP pourrait également servir de **relais** entre les commune et le niveau régional. De par son enclavage en territoire namurois et sa proximité avec les communes, le BEP permettrait que toute commune, indépendamment de leur taille, puisse bénéficier d'une offre sur l'ouverture des données. L'objectif serait que les données sur la plateforme du BEP bénéficient d'une visibilité au niveau régional (avec Digital Wallonia). L'étude de la Commission Européenne sur les portails Open Data mentionne clairement que

les initiatives non intégrées à un niveau supra réduisent la publicité et donc la potentielle réutilisation de ces données : l'étude mentionne que « seulement 13% des pays ont intégrés tous les portails régionaux dans le portail national » [8].

Le BEP est une organisation qui a développé, au cours de ces dernières années, des services de support de plus en plus importants et efficaces (ressources humaines, finances et comptabilités, IT, ...). Ces services de support sont majoritairement destinés à la gestion interne du BEP. Néanmoins, les communes viennent également trouver le BEP pour avoir un conseil dans les bonnes pratiques de la gestion d'un service de support. Ces services de support ainsi que les départements spécifiques du BEP pourraient être sollicités dans la mission Open Data du BEP envers les communes. En effet, les différentes entités du BEP sont compétents pour évaluer le type de données qui existent et qui peuvent être facilement ouvertes. En fonction des données à ouvrir, les départements adéquats pourraient être mobilisés pour assurer une communication entre les personnes du BEP proposant un service d'ouverture des données et la commune disposée à partager les données. Dans cette dynamique – et on revient à un point précédemment cité -, le BEP aurait un grand intérêt à montrer, une fois de plus, l'exemple en partageant les données de ses différents départements : l'efficacité de futures démarches ne serait que meilleure lors de missions ultérieures avec les communes.

3 Open Data

3.1 Plateformes Open Data

3.1.1 Vision du BEP sur une plateforme Open Data

Avant d'aborder les besoins du BEP en matière de plateforme Open Data, il est utile de préciser quels sont les objectifs poursuivis par le Bureau Économique de la Province de Namur. Ces objectifs stratégiques permettront de lister et de prioriser les besoins du BEP.

1. Le premier but du BEP est d'apporter une transparence sur l'organisation BEP ainsi que sur les missions qui lui sont confiées. La transparence est donc animée par l'objectif d'améliorer l'accessibilité à l'information.
2. Le BEP veut rendre la ville plus intelligente dans sa globalité en tenant compte de ses enjeux (économie, mobilité, environnement (comptabilité énergétique automatisée), citoyenneté, gestion des déchets, sécurité, démographie, ...). Il veut améliorer les services rendus aux citoyens grâce à de l'information de qualité et grâce à de l'information en temps réel [4].
3. Le BEP a pour mission de sensibiliser les citoyens et les communes à l'extraction de la connaissance venant de l'Open Data dans l'optique d'avoir des citoyens qualifiés [4]. L'engagement du citoyen est par conséquent un but poursuivi.
4. Le BEP se veut être une entreprise efficace et il recherche également cette efficacité dans les solutions et les outils qu'ils proposent. De plus, le BEP prône l'autonomie de ses collaborateurs grâce aux outils informatiques adéquats : les besoins de la plateforme seront également évalués, du point de vue des collaborateurs en charge de celle-ci, vis-à-vis de sa facilité d'utilisation, son efficacité et de l'autonomie qu'elle confèrerait à ses administrateurs.

3.1.2 Analyse des besoins

3.1.2.1 Dynamique de la récolte des besoins

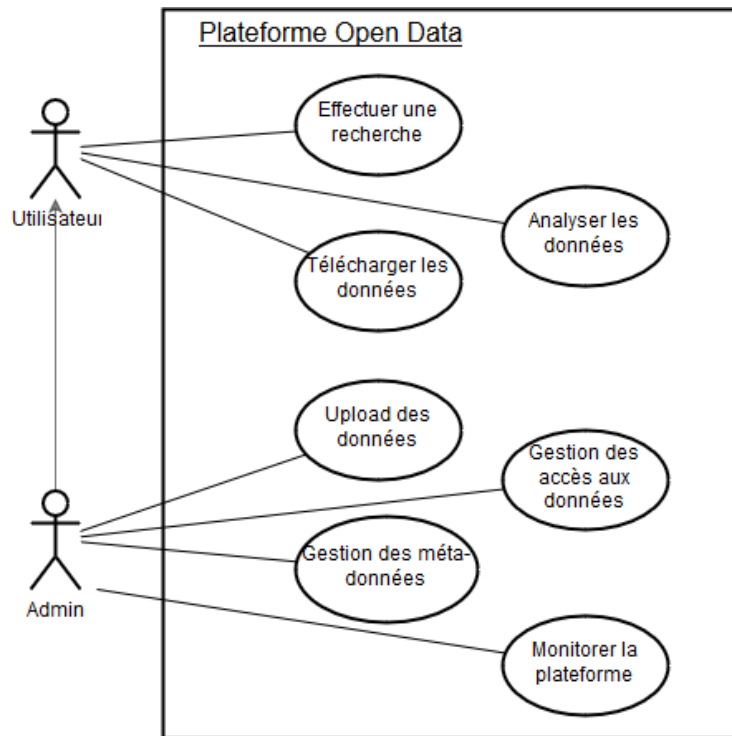
Comment va se dérouler l'élucidation des besoins ? Dans un premier temps, je me suis informé sur ce qu'une plateforme Open Data devait contenir pour maximiser son succès et donc la réutilisation des données : j'ai donc focalisé sur le « domain understanding ». Ensuite, nous avons rencontré un provider de plateforme Open Data, l'entreprise OpenDataSoft : cela a permis de, manière informelle, de voir ce que le BEP avait comme besoin (besoins fonctionnels et non-fonctionnels) pour une future plateforme.

En groupant ces 2 sources d'informations, nous avons pu créer une liste exhaustive de critères ainsi que leur importance pour évaluer et comparer différentes plateformes Open Data : cette liste exhaustive de besoins fût présentée aux personnes du BEP impliquées donc le projet pour validation, ajout ou suppression. Dans un premier temps, la liste des besoins sera énumérée dans la section « 3.1.2.2 Identification des besoins ». Ensuite, les différentes plateformes seront analysées dans la partie « 3.1.3 Analyse des plateformes » au regard de chaque critère. Finalement, une plateforme sera recommandée sur base de leur analyse, dans la section « 3.1.3 Analyse des plateformes ».

3.1.2.2 Identification des besoins

3.1.2.2.1 Besoins fonctionnels

Les besoins fonctionnels principaux vont d'abord être présentés sous la forme de Use Cases pour avoir une vue d'ensemble de ce que l'utilisateur ainsi que l'administrateur attendent d'une plateforme Open Data. Ensuite, les use cases seront détaillés dans les sections suivantes.



3.1.2.2.1.1 Must have

Accessibilité des données via le Web

Il est indispensable que notre plateforme Open Data soit disponible sur le Web. En effet, c'est le moyen de communication qui fera en sorte qu'un maximum de personnes ait la possibilité d'accéder aux données (**accessible**). De plus, l'utilisateur n'a besoin d'aucune inscription que pour effectuer les actions de base sur la plateforme : recherche de données, téléchargement de données, visualisation des données,

Plateforme et données facilement recherchables

La plateforme permet d'effectuer des recherches sur base :

- de texte. Les recherches sont effectuées sur les noms de données ainsi que sur les meta-données que les jeux de données possèdent.
- des meta-données liées aux données. Les recherches effectuées peuvent être réalisées sur la date de création, sur la date de mise à jour, sur le thème des données, sur le département/organisation ayant mis à disposition les données, ...
- de tri sur les données par : le nombre de vues, le nombre de téléchargement, date de mise à jour des données, ordre alphabétique, ...

La plateforme fournit la possibilité de taguer des jeux de données avec des mots-clés afin d'améliorer les recherches faites par les utilisateurs de la plateforme.

Téléchargement des données

Le format des données, si on veut atteindre le niveau 3 du « 5 star model » de Tim Berners-Lee, doit contenir au moins un format **machine-readable** et **non-propriétaire**. Nous voulons que ce format de

données soit également accessible aux non-techniciens. Nous estimons donc que les formats répondants à ces exigences sont le csv (comma separated values) ou tsv (tab separated values) et le JSON.

A côté de ces formats non propriétaires, il est toujours utile, pour la facilité, d'avoir un format de fichier tel que Excel (xls etxlsx) et OpenDocument Spreadsheet (.ods) quand on veut jouer rapidement avec les données via un tableau ou graphique dynamique croisé.

De plus, les données doivent pouvoir être récupérées via une API (Application Program Interface) au format JSON. Cela sera utilisé par ceux qui développent des applications dans l'optique de mettre à jour régulièrement les données dans leur application. L'API doit être suffisamment flexible que pour :

- appliquer des filtres ;
- extraire seulement les champs nécessaires ;
- (optionnel) trier les données envoyées ;
- (optionnel) faire une agrégation de type « group by ».

Analyse de données : affichage des données brutes, tris, simples filtres et fonctionnalités avancées (graphiques, cartes, ...)

L'utilisateur peut visionner les données directement sur la plateforme, sans avoir à les télécharger : il se fait ainsi une idée du genre de données à sa disposition ainsi que les valeurs prises par certaines instances du jeu de données. Cette fonctionnalité doit également supporter le tri ainsi que de simples filtres.

L'utilisateur a la possibilité de manipuler des graphiques préétablis et/ou de créer ses propres graphiques de toutes pièces. L'utilisateur peut également réaliser des filtres sur les données entrantes dans le graphique.

Voici les graphiques de bases qu'un utilisateur pourrait avoir besoin :

- L'utilisateur peut visualiser les données sur une carte, il y a des infos-bulles affichant davantage d'informations concernant une instance spécifique sur la carte.
- S'il réalise des graphiques de type « bar chart », il peut utiliser les agrégateurs suivants : comptage, somme, moyenne et pourcentage (basé sur la somme ou sur le comptage). La plateforme idéale serait capable de tenir compte jusqu'à 2 variables de ventilations : le graphique afficherait dans ce cas la valeur de l'agrégation pour chaque combinaison de valeurs rencontrées dans les données.
- S'il veut observer la relation entre 2 variables quantitatives, l'utilisateur peut faire un graphique de type « scatter plot ».
- S'il veut voir comment est distribuée une variable continue, il peut réaliser un histogramme et, si une variable de ventilation est choisie, un graphique de type « box plot ».
- Des graphiques représentant des observations au cours du temps peuvent également être construits (line chart).

La plateforme Open Data idéale (*soyons un peu utopiste*) serait celle pouvant proposer des graphiques sur base des champs dont l'utilisateur aurait besoin et sur base des meta-données : filtres sous forme d'un calendrier pour les dates, les variables de ventilations pourraient être différenciées des variables d'analyse, les agrégations de variables seraient vérifiées en conformité avec leur meta-données. Cela

aurait pour objectif d'indiquer des analyses possibles à l'utilisateur tout en lui retirant la possibilité de faire des analyses n'ayant aucun sens.

Upload de données : possibilités ETL avancées

La plateforme Open Data permet d'avoir des fonctionnalités ETL (Extract-Transform-Load) avancées qui permettront de pouvoir fournir un maximum de données, avec la meilleure qualité possible et grâce à des processus automatisés. Les triggers pour lancer un processus sont basés sur le temps (« *time event trigger* ») ou sur la présence d'un nouveau fichier, d'un fichier rafraîchi ou alors d'un folder (« *file event trigger* ») ou encore sur la date de rafraîchissement d'un jeu de données présents sur la plateforme.

- **Extract:** Les données alimentant la plateforme peuvent être lues à partir d'un maximum de types de sources :
 - o fichiers plats (csv, tsv, JSON, ...) ;
 - o données provenant d'une API ;
 - o fichier Excel et OpenDocument Spreadsheet ;
 - o accès en lecture dans des bases de données (prise en compte de l'environnement technique du BEP) : SQL Server, bases de données géographiques de la solution ESRI.
- **Transform:** la plateforme permet de réaliser différentes transformations aux données afin d'augmenter la potentiel résidant au sein de la données. Les fonctionnalités attendues sont les suivantes:
 - o cleaning,
 - o parsing,
 - o vérification sur la qualité de données,
 - o jointure, enrichissement de données
 - o transformation de codes géographiques.
- **Load:** la plateforme supporte aussi bien le chargement incrémental (incremental load, avec une date de chargement des records concernés) et le chargement complet (full load). La plateforme propose-t-elle une solution de versionning pour les dataset qui sont écrasés à chaque fois ?

La plateforme doit guider le gestionnaire de celle-ci grâce à un ETL intuitif, puissant et performant : le but est de faciliter le travail du gestionnaire ainsi que le rendre efficace.

Des données peuvent également être mises à jour en « near real-time » : c'est-à-dire qu'il doit être possible de mettre à jour des données toutes les 10 minutes par exemple. Cet aspect est guidé par la volonté de s'inscrire dans des projets SmartCity où des capteurs récolteraient diverses informations et ces dernières seraient envoyées vers la plateforme Open Data.

Gestion des accès aux données et à la plateforme

La plateforme doit offrir la possibilité de pouvoir gérer la publication totale ou partielle de données. Par « partielle », nous entendons que certaines lignes et/ou certaines colonnes du jeu de données peuvent être rendues non visibles pour les utilisateurs finaux. Il est également possible d'héberger des sets de données sans qu'ils soient visibles pour les utilisateurs finaux.

Les accès aux données peuvent être définis par groupe d'utilisateurs ainsi que par utilisateurs précis : si cette option est possible, nous acceptons que la plateforme requière une inscription pour la gestion d'utilisateurs ou groupes d'utilisateurs particuliers.

La plateforme idéale posséderait des processus permettant la validation (workflow de validation) et la publication de datasets. Exemple : une personne de la comptabilité validerait les chiffres et les rendrait par la même occasion visible de tous.

La plateforme doit pouvoir donner des droits spécifiques à des data holder pour le chargement les données sur la plateforme. En effet, à moyen et long terme, si davantage de ressources sont nécessaires ou si il y a un changement de rôles au sein des parties prenantes, il sera nécessaire de pouvoir facilement donner accès à la plateforme dans le but de mettre davantage de données à disposition.

Création et gestion de meta-données

La plateforme doit donner la possibilité de créer et de tenir à jour les meta-données relatives aux données. Au plus l'encodage de meta-données est minimisé, au mieux cela est : cela signifie que certaines meta-données doivent être rafraichies/calculée automatiquement comme, par exemple, la date de création des données, la date de mises à jour des données, la fréquence de rafraichissement des données, la qualité des données, ...

Les meta-données doivent suivre un format standard : DCAT [14]. De plus, la plateforme doit permettre d'enregistrer au minimum les meta-data suivantes (le nom de la propriété lié au format DCAT est présent entre parenthèses) :

- Nom du dataset (« *title* »)
- Description du dataset (« *description* »)
- Date de publication du dataset = date de mise à disposition du jeu de données sur la plateforme (« *issued* »)
- Date de modification du dataset (« *modified* »)
- Langue du dataset (« *language* »)
- Entité responsable de la mise à disposition des données (« *publisher* »)
- Fréquence de mise à jour des données (« *accrualPeriodicity* »)
- Identifiant unique du dataset (« *identifier* »)
- Zone couverte par le jeu de données (« *spatial* »)
- Période temporelle couverte par le jeu de données (« *temporal* »)
- Sujet principal des données (« *theme* »)
- Mot clé sur le dataset (« *keyword* »)
- Licence du dataset (« *license* »)

Avoir une vue d'ensemble sur les meta-données est potentiellement utile pour ceux qui administrent la plateforme et pour les plateformes qui voudraient rechercher automatiquement le contenu de ces meta-données dans l'optique d'un référencement à un niveau supra. Toutes les meta-données doivent pouvoir également être facilement à disposition via les formats de données précédemment cités (csv ou tsv, Excel, OpenDocument Spreadsheet, JSON (via API)).

Appliquer une charte graphique à la plateforme

La plateforme Open Data doit suivre la charte graphique du Bureau Économique de la Province de Namur : que ce soit sur base de développement sur mesure ou alors sur base de fonctionnalités prévues à cet effet.

Si la plateforme peut être facilement modifiée, alors ce sera un plus pour la plateforme. Les fonctionnalités suivantes seraient donc recherchées : les éléments de la page Web peuvent facilement être déplacés, le thème peut être modifié, un logo de l'institution peut être ajouté et être présent sur toutes les pages de navigation. Pour d'éventuels besoins de développement sur mesure, il est facile d'intégrer des scripts HTML et CSS pour customiser le look des pages prédéfinies.

3.1.2.2.1.2 Should have

Faire le suivi du chargement et de l'utilisation des datasets et de l'API

Les gestionnaires de la plateforme peuvent monitorer la plateforme pour :

- Faire le suivi de la mise à jour des datasets :
 - o vérification si les jobs schedulés se sont correctement exécutés (logging et overview si des erreurs dans les jobs se sont manifestées ainsi que le temps d'exécution du job),
 - o monitoring de l'insertion des nouvelles données lorsque c'est un chargement incrémental.
- Faire le suivi sur la réutilisation des données : par dataset, catégories de datasets, jours (semaine, mois, année), type de réutilisation (graphique, download à la main, API), lieu du réutilisateur, ...

Alimenter un portail à un niveau supra

Les informations présentes (données et méta-données) sur la plateforme peuvent être récoltées automatiquement par une plateforme à un niveau supra. Le but serait que la plateforme Open Data du BEP puisse recevoir une publicité plus large grâce à des plateformes Open Data dont la couverture territoriale est plus large.

3.1.2.2.1.3 Could have

Interaction de la communauté autour de la donnée

La plateforme permet de recevoir des feedbacks, via un formulaire en ligne, des utilisateurs où ils peuvent commenter la qualité des données, demander davantage d'informations sur les données présentes, signaler les données en erreur, demander de mettre à disposition de nouveaux jeux de données, ... À titre d'exemples, il serait intéressant que les utilisateurs :

- puissent donner une note quant à leur satisfaction par rapport à un jeu de données spécifique ;
- soient informés lorsque des jeux de données sont mis à jour ou une organisation auquel un utilisateur s'est abonné a publié un nouveau dataset ;
- puissent rapidement partager des datasets et leur impression au sujet de ceux-ci sur les réseaux sociaux.

Pour aller encore plus loin, la plateforme pourrait permettre aux utilisateurs de publier les données qu'ils ont modifiées car ils y ont apporté une valeur ajoutée : combinaison de jeux de données, nettoyage, analyses intéressantes et uniques, ... Dans la continuité de la réflexion, il pourrait être intéressant de relier les données avec les réutilisations et les applications ayant utilisés ces données.

Dans l'optique de la communauté, il pourrait être intéressant que la plateforme ait un widget permettant d'afficher les tweets relatifs à une channel Twitter ou à un hashtag particulier. Avoir des fonctionnalités qui améliorent le processus de sélection de données ainsi que leur diffusion est un élément important

pour assurer l'effervescence et l'utilité autour d'une plateforme Open Data : toute autre idée proposée par la plateforme sera la bienvenue pour mener à la réalisation de cet objectif.

Enregistrer/exporter les visualisations réalisées par l'utilisateur

Si la plateforme permet de réaliser des visualisations, il serait alors intéressant que l'utilisateur puisse les sauvegarder (dans un éventuel espace personnel) et/ou les exporter en divers formats (jpg, png, svg, ...).

Support technique sur la plateforme pour les utilisateurs (tutoriels, ...)

La priorité est que la plateforme Open Data soit intuitive et claire pour les personnes qui vont surfer sur les différentes pages de celle-ci. Mais il sera également évalué si la plateforme a une série de tutoriels (vidéos, documentation grand public, capture d'écran, ...) permettant de donner des conseils, des astuces pour les utilisateurs et les administrateurs de la plateforme.

Possibilité que le portail puisse extraire les meta-données sur un autre portail

Dans l'éventualité où le BEP devrait centraliser les données de portails Open Data à un niveau communal et que ces portails Open Data existent, il serait dès lors utile que cette fonctionnalité soit déjà prévue dans la plateforme.

3.1.2.2.2 Besoins non fonctionnels

« Look and feel » de la plateforme

Les interfaces pour l'utilisateur, dans sa recherche, et pour l'administrateur, dans la gestion opérationnelle de la plateforme, sont claires et n'amènent pas de confusion. De plus, les interfaces pour les utilisateurs sont compréhensibles par des non-techniciens : la navigation doit être facile et intuitive. Il ne doit pas y avoir de friction lorsque l'utilisateur navigue sur la plateforme Open Data.

Evaluation du lock-in

La plateforme doit présenter un lock-in relativement faible. On évaluera dans quelle mesure le BEP peut sortir de la solution choisie dans l'optique d'aller vers une autre solution : il sera donc question de regarder quels sont les moyens considérés pour sortir de la plateforme :

- Extraction en bulk de toutes les données et méta-données ;
- Format dans lequel les informations seront extraites.

Disponibilité

La disponibilité de la plateforme doit être au minimum de 99,5% du temps.

Capacité

Il faudra considérer quelle capacité la plateforme peut délivrer lorsque les utilisateurs naviguent dessus, téléchargent des données ou encore font des appels API.

Sécurité

Les requêtes générées par les utilisateurs vers la plateforme sont encryptées : les données envoyées de la plateforme vers l'utilisateur ne peuvent pas être lues ni altérées par une tierce personne.

Conservation des données (back-up)

La plateforme offre des stratégies pour la pérennité des données et meta-données présentes sur le portail Open Data.

3.1.3 Analyse des plateformes

3.1.3.1 Contexte et stratégie IT au sein du BEP

Premièrement, il est utile d'analyser quelles sont les stratégies opérées par le département IT au cours de ces dernières années : cela situera le contexte et la philosophie dans lesquels la plateforme Open Data devra s'intégrer. La façon de travailler au sein de l'IT, comme beaucoup d'entreprises, a évolué en fonction des tendances technologiques apparaissant sur le marché.

Historiquement, le département IT réalisait ou faisait réaliser des solutions sur mesure répondant aux besoins du business mais la maintenance ainsi que l'évolutivité de ces solutions étaient difficiles à garantir. De manière générale, le paysage informatique a fortement évolué. Le nombre d'acteurs a fortement augmenté dû à la façon de proposer les logiciels : on est passé des licences – où les solutions logiciels sont installés sur les propres serveurs des entreprises – aux logiciels proposés dans le cloud.

Etant donné cette évolution majeure, la stratégie IT du BEP s'est adaptée. L'équipe IT du BEP a fait le tour des différents départements pour lister quels sont les différents besoins présents. Ces besoins ont été analysés pour en retirer leur essence propre et avoir une liste de besoins standards pouvant être rencontrés par de multiples départements : l'idée est de dégager ce qui est transversal.

Le fait même de vouloir répondre à un besoin primaire implique par conséquent de regarder à des solutions standards : pour chaque besoin, une solution standard va être recherchée. Fort de ces tendances et de la manière actuelle de travailler au BEP, la ligne stratégique de l'IT va prendre les concrétisations suivantes :

- Diminution des développements en interne et/ou des développements sur mesure ;
- « Un outil pour un besoin » ;
- Interopérabilité des outils/solutions ;
- Choix tournés davantage vers le cloud (diminution des coûts et gain en flexibilité) [4] ;
- Volonté d'avoir des processus transparents.

3.1.3.2 Approche utilisée pour l'analyse des plateformes

Chaque plateforme sera analysée en fonction des besoins fonctionnels précédemment développés. Les plateformes ne seront pas évaluées en fonction des besoins non fonctionnels (excepté le « look and feel ») : les informations nécessaires peuvent exiger une connaissance trop technique que je n'ai pas. Je me concentre sur les fonctionnalités que chaque plateforme propose.

Après un tour rapide du marché, nous avons identifié les principaux acteurs/solutions proposant une plateforme Open Data. 2 solutions sont des solutions propriétaires (OpenDataSoft et Socrata) et la dernière est CKAN. Ce sont les solutions les plus utilisées et couvrent une très grande partie du marché des plateformes Open Data. De plus, ce choix permet aussi de voir ce que propose des solutions propriétaires « clés sur porte » (Socrata et OpenDataSoft) en comparaison avec une alternative en open source (CKAN).

Chaque plateforme analysée suivra la structure décrite ci-dessous :

- Description générale
- Besoins fonctionnels
- Tarification (si possibilité d'obtenir)

3.1.3.2.1 Analyse préliminaire sur CKAN

CKAN (Comprehensive Knowledge Archive Network) est un portail open source permettant de publier, de partager, de chercher et d'utiliser des jeux des données. La plateforme possède un grand nombre d'extensions (plus de 300) qui permettent de rajouter des fonctionnalités au portail. Le back-end est en Python tandis que le front-end est en HTML et Javascript.

Les plus grands portails Open Data existants actuellement sont supporté par cette solution : gouvernement des Etats-Unis d'Amérique, gouvernement de Grande Bretagne, ... Les grandes institutions ont la capacité et les connaissances pour héberger la solution sur leur propre serveur tandis que les plus petites administrations peuvent héberger ces solutions dans le cloud, en faisant appel à une entreprise tierce pour les aider.

En visitant quelques portails et en discutant avec certains administrateurs, chacun portail a sa variante dû aux nombreuses extensions possibles et à la possibilité aussi d'aller modifier le code source.

Plusieurs plateformes Open Data existantes basées sur CKAN ont été analysées. On a pu remarquer certaines limitations ou faiblesses lié à cette technologie :

- Cette technologie nécessiterait du sur-mesure et des compétences en HTML et Python : la stratégie IT du BEP n'est plus à faire du sur-mesure et ils ne sont pas prêts à investir pour du personnel qualifié en HTML ou Python.
- CKAN semble encore fort attaché à son origine : c'est-à-dire une collection de fichiers mises à disposition sans réelle valorisation des données. A sachant que l'un des objectifs principaux est la transparence, choisir CKAN n'aiderait à la réalisation de cet objectif.
- Les fonctionnalités recherchées pour une plateforme Open Data sont moins abouties et moins « user-friendly » :
 - o Les utilisateurs de la plateforme seront moins engagés et la plateforme ne rencontrerait pas l'intérêt du public :
 - notre analyse sur CKAN n'a pas révélé la possibilité de faire des analyses de données au sein même de la plateforme ;
 - les formats de téléchargements varient d'un dataset à l'autre sans qu'on en comprenne véritablement la raison ;
 - le « look & feel » de la plateforme n'est pas très attirant et l'interface de navigation n'est pas intuitive pour les utilisateurs ;
 - un dataset déterminé peut comprendre de multiples fichiers (un fichier par année alors que les différents fichiers pourraient ne constituer qu'un dataset en les « empilant ») ;
 - les développeurs d'application n'auront pas un support suffisant et clair que pour faire des appels API.

En analysant les plateformes avec une large visibilité (celles avec potentiellement le plus de moyens) on remarque qu'il y a un manque de standardisation dans les formats des datasets (certains datasets ne sont présents que sous un format propriétaire ou sous un format non prévu pour l'analyse de donnée) et/ou dans le nombre de datasets (un fichier par année par exemple). On voit donc bien que la plateforme CKAN n'est pas un outil de standardisation de ce qui entre sur la plateforme : il n'offre pas des routines ETL pour mettre jour les données.

Si la standardisation ne peut être réalisée grâce à la plateforme, alors cela doit se dérouler en amont de la plateforme. Vu que le BEP ne bénéficie pas des outils pour faire cela en amont, le BEP attend de la plateforme qu'elle offre les outils pour les bases de l'ETL : il est donc primordial que l'outil en question puisse répondre à cet objectif afin de s'assurer que les données soient standardisées.

Sur base de cette étude préliminaire, CKAN ne sera pas davantage analysé car les besoins primordiaux que le BEP attend de la plateforme ne sont pas rencontrés. CKAN n'est pas un candidat qui répondra aux attentes du BEP.

3.1.3.3 Socrata

3.1.3.3.1 Description générale

Socrata est une entreprise américaine fondée en 2007. Elle propose une plateforme Open Data en SaaS (Solution as a Service). La majorité de ses clients se situe principalement aux Etats-Unis et ce sont des clients publics. Ils vendent majoritairement leur solution à des villes, des comtés ainsi que des organisations à un niveau fédéral.

La façon dont il se présente est d'aider les organisations à rationaliser les processus et rendre l'administration plus efficace. Il y a également une volonté d'ouvrir un maximum les données afin que des tierces – et ce sont principalement les développeurs d'application qui sont visés – puissent bénéficier de ces données pour créer de nouveaux services. Enfin, ils jouent aussi sur la volonté de transparence des organisations publiques.

3.1.3.3.2 Besoins fonctionnels

3.1.3.3.2.1 Must have

Accessibilité des données via le Web

Socrata est une plateforme proposée en SaaS : la consultation (pour les utilisateurs finaux) ainsi que l'administration de la plateforme se fait en ligne – simplement grâce à un navigateur et une connexion internet. Les administrateurs se loggent sur la plateforme pour charger et maintenir à jour les datasets dont ils sont responsables.

Plateforme et données facilement recherchables

La recherche sur la plateforme Socrata est relativement efficace car c'est une série de meta-données qui sont indexées [15]. Il y a aussi le principe de stemming présent : des mots ayant une base commune sont proposés dans les résultats de recherche. Néanmoins, le français n'est que partiellement supporté par la plateforme actuellement : ce qui réduit probablement l'efficacité du stemming. Les jeux de données peuvent être filtrés par thème, par les types de vues (calendrier, graphiques, Data Lens, filtered view, ...), par producteur de données, par mots clés. Contrairement à OpenDataSoft, si une valeur est sélectionnée, toutes les autres valeurs du même type de filtre peuvent également être sélectionnées. Mais

contrairement à OpenDataSoft, il n'est pas possible d'avoir une vue sur le nombre de datasets concerné par chacune des valeurs des filtres.

Téléchargement des données

Les différents formats pour télécharger les données sont les suivants : CSV, CSV pour Excel, CSV pour Excel (Europe), tsv pour Excel, JSON, RDF, RSS, XML, KML, KMZ, Shapefile (pas de limitation sur le nombre de records contrairement à OpenDataSoft), GeoJSON. Si le jeu de données contient trop de données pour un utilisateur, il peut effectuer des filtres sur celui-ci et n'exporter que le jeu de données filtré. Une API simple d'utilisation donne une très grande valeur ajoutée à la plateforme Socrata : il n'est pas étonnant que les données soient autant demandées sur les plateformes hébergées par Socrata. L'API de Socrata est très puissante et simple d'utilisation : elle repose sur le langage SQL [16]. Donc même pour les non techniciens, ils peuvent s'en sortir sans trop de difficultés pour effectuer des filtres, la sélection de certains champs, les agrégations, et le triage. De plus, l'API de Socrata permet également de choisir le format des données (json, xml et csv) lorsqu'on veut obtenir les données via l'API. Si une application réalise souvent des appels API, Socrata peut bloquer certaines de ces requêtes en se basant sur l'adresse IP (Socrata effectue cela pour qu'il y ait un partage équitable entre les utilisateurs et qu'aucun utilisateur monopolise la plateforme) : pour éviter cette limitation, il est conseillé que l'application demande un token à la plateforme ainsi la limite du nombre de requêtes API sera plus élevée.

Analyse de données : affichage des données brutes, tris, simples filtres et fonctionnalités avancées (graphiques, cartes, ...)

Socrata offre également une plateforme où il est également possible d'observer directement les données dans une table et de réaliser des graphiques, des cartes. Mais l'utilisation de ces dernières fonctionnalités n'est pas très intuitive et ne permet pas, par exemple, de compter le nombre de fois qu'une valeur est prise par un champ.

Upload de données : possibilités ETL avancées

Concernant l'extraction, les transformations et le rafraichissement des données, il est clair que Socrata n'excelle pas dans ce domaine. En effet, il est majoritairement focalisé sur la réutilisation des données.

- Socrata possède au sein de sa plateforme un outil ('Append and Replace Wizard') qui
 - o permet l'extraction de fichiers sous un grand nombre de formats (flat files, URL, Excel, ...);
 - o ne permet aucune transformation : pas de filtres, de géo-coding d'adresses, de jointures, du parsing de champs texte ;
 - o ne permet pas le scheduling de job pour mettre à jour les datasets ;
 - o n'a pas de taille limite concernant la lecture des fichiers sources ;
 - o supporte le full load (chaque dataset, indépendamment de la méthode de chargement, peut avoir un versionning, ce qui est intéressant pour avoir les différentes photos d'un même dataset) et l'insert ;
 - o ne supporte pas l'update ou le delete de records existants.
- Socrata propose des alternatives en fonction du contexte du client :

- si le client veut juste automatiser la lecture et le rafraichissement de données, alors il peut opter pour Socrata DataSync qui est une application Java gratuite qui doit être installée chez le client (et maintenue par celui-ci).
 - Il doit utiliser en fait le 'Windows Task Manager' (Windows) ou le 'Cron' (Linux) pour réaliser le scheduling d'un job.
 - De plus, il est préférable d'avoir également un token pour ne pas que les requêtes envoyées par le software 'Socrata DataSync' ne soient bloquées par la plateforme Socrata (principe de throttling utilisé pour garantir un partage équitable des ressources entre les différents utilisateurs de la plateforme).
 - L'outil supporte toutes les méthodes de chargement de données : replace, update, insert et delete.
- si le client possède déjà un outil ETL ou un outil s'en rapprochant ou si le client ne possède aucun outil de ce type, alors il propose
 - d'utiliser Safe FME's mais c'est outil est payant.
 - si le client utilise Pentaho Kettle (gratuit), alors il peut télécharger un plugin qui permet de rafraichir les datasets sur la plateforme Socrata :
 - le client doit également télécharger Socrata DataSync car le plugin fait appel à ce software ;
 - le client ne doit pas oublier de demander également un token.
 - si le client utilise R (gratuit), alors il peut télécharger le package 'RSocrata' pour charger les données sur la plateforme. Ici à nouveau, il sera requis d'automatiser un job (écrit en R) avec le 'Windows Task Scheduler'.
 - si le client utilise d'autres outils, alors Socrata propose d'utiliser l'API de Socrata pour rafraichir les données sur Socrata : il est également possible de charger les méta-données via l'API.

Gestion des accès aux données et à la plateforme

Concernant la gestion de l'affichage des datasets, soit le dataset est public soit le dataset est privé et alors dans ce cas, seulement les utilisateurs autorisés peuvent consulter le dataset. Gérer des champs et/ou des records doivent être gérés dans la partie ETL mais vu que Socrata ne contient pas ce type de fonctionnalités, c'est alors l'outil ETL servant à préparer les datasets qui doit s'occuper de cette partie. Socrata ne possède pas de workflow de validation : néanmoins, un dataset peut être gardé privé et partagés entre certains utilisateurs et dès que les données sont validées, un utilisateur peut mettre les données à disposition du public. La plateforme permet de gérer facilement les différents droits aux utilisateurs : il est facilement concevable que des droits soient attribués à plusieurs personnes pour charger et administrer les données sur la plateforme Open Data.

Création et gestion de meta-données

Socrata suit les standard en termes de métadonnées : il utilise le standard DCAT.

Appliquer une charte graphique à la plateforme

L'apparence de la plateforme peut être gérée par l'administrateur mais il n'est pas très clair quelles sont les possibilités et les limites exactes.

3.1.3.3.2.2 Should have

Faire le suivi du chargement et de l'utilisation des datasets et de l'API

Concernant le monitoring de la plateforme au niveau du chargement des données, tout cela est géré au sein de la plateforme et cela est possible quelque que soit la méthode de chargement choisie (Wizard, Socrata DataSync, API Publisher de Socrata) pour mettre à jours les datasets. Et au niveau de l'utilisation de la plateforme, une série de statistiques sont récoltées et sont affichées aux administrateurs de la plateforme qui peuvent par la même occasion comprendre l'utilisation faite par les utilisateurs.

Alimenter un portail à un niveau supra

Les meta-données de Socrata peuvent être récoltées par un autre portail Open Data car elles peuvent être extraites grâce à son API pointant vers son catalogue de jeux de données.

3.1.3.3.2.3 Could have

Interaction de la communauté autour de la donnée

Socrata donne la possibilité de discuter autour d'un dataset et/ou de contacter directement le propriétaire du dataset. Il est même possible que l'utilisateur donne des étoiles aux datasets pour dire sa satisfaction par rapport aux données qu'il utilise. Socrata permet également de mettre en avant de nouveaux datasets et des datasets méconnus en les proposant dans les résultats de recherche de type « features contents » (ce sont des datasets mis en avant lors d'une recherche effectuée par ceux qui surfent). Socrata offre aussi l'occasion que les utilisateurs partagent les données via les réseaux sociaux (Twitter, Facebook, Youtube, LinkedIn, Github).

Enregistrer/exporter les visualisations réalisées par l'utilisateur

Les graphiques et les tableaux affichés à l'écran peuvent être facilement réintégrés dans d'autres sites web en faisant simplement un copier-coller du code HTML généré.

Support technique sur la plateforme pour les utilisateurs (tutoriels, ...)

Socrata offre pas mal de support qui se révèle plutôt technique : il y a peu de support pour la vulgarisation et la conscientisation à l'outil et à l'Open Data. Ils sont fortement tournés vers le côté technique (API, rafraichissement des données, ...).

Possibilité que le portail puisse extraire les meta-données sur un autre portail

Avec les dernières releases de la plateformes, il est désormais possible que Socrata puissent puiser dans divers catalogues de méta-données (ArcGIS et tous portails Open Data avec une extension « data.json ») [17].

3.1.3.3.3 Tarification

Il a été difficile d'obtenir une tarification précise mais d'après le représentant Socrata que j'ai contacté et sur base de nos besoins, il estime qu'on est aux alentours de 12 000 € par an.

Néanmoins, je recommande d'envoyer un cahier des charges à Socrata pour qu'il effectue une estimation de l'abonnement annuel. En effet, la tarification se base sur des modules. Il faudrait donc pouvoir estimer quels sont les modules nécessaires à la plateforme. La tarification chez Socrata ne dépend ni du nombre de datasets ni de volumes de ceux-ci et le représentant m'a assuré qu'aucun des clients qu'il connaît pour

le moment a rencontré un problème dans le nombre de datasets, dans le nombres d'appels fait à l'API ou dans la taille des jeux de données présents sur la plateforme.

3.1.3.4 *OpenDataSoft*

3.1.3.4.1 Description générale

OpenDataSoft est une entreprise française existante depuis 2011. Il propose une solution sous forme SaaS (Software as a Service) où le back end est en Python et le front end est en HTML et Angular JS. OpenDataSoft insiste sur une fonctionnalité de recherche efficace grâce à ElasticSearch, une méthode permettant d'indexer efficacement un ensemble de données. Le client peut choisir d'héberger la solution chez Amazon (Amazon Web Services) ou chez Microsoft (Microsoft Azure).

La majorité de ses clients se situe en France mais ils sont également sur le marché américain où ils comptent déjà quelques villes (Durham, Cary) au sein de leur clientèle. Les clients d'OpenDataSoft sont issus, pour 70% du secteur public et des organisations sans but lucratif, mais ils sont également présents dans des entreprises privées (EDF, Total, Suez, Engie) pour fournir des solutions reposant sur les technologies plus ou moins proches de la plateforme Open Data qu'il propose.

OpenDataSoft se propose comme une solution « clé sur porte » où toutes les fonctionnalités sont présentes dans la plateforme (voir section « Tarification ») et où la facilité d'utilisation est présente aussi bien au niveau des utilisateurs finaux (ceux qui surfent sur la plateforme) qu'au niveau de ceux qui seront en charge d'administrer la plateforme. OpenDataSoft s'adresse davantage au grand public que Socrata.

3.1.3.4.2 Besoins fonctionnels

3.1.3.4.2.1 Must have

Accessibilité des données via le Web

OpenDataSoft est une plateforme proposée en SaaS présente dans le cloud : son utilisation et son administration se gère directement depuis un navigateur Internet. Les administrateurs se loggent sur la plateforme : ils peuvent y charger et maintenir à jour les datasets dont ils sont responsables.

Plateforme et données facilement recherchables

OpenDataSoft propose une plateforme permettant d'effectuer des recherches : les recherches va rechercher dans les métadonnées. Indépendamment de la recherche textuelle, l'utilisateur peut également filtrer par vues (ce sont les différents widgets présents pour un jeu de données (Analyse, Carte, Image, ...)), par année de modification, par producteur, par mot-clés, par thème, par langue, par zone géographique (il suffit aux administrateurs de l'activer). Malheureusement une fois une valeur sélectionnée pour un type de filtre, il n'est pas possible de choisir de filtrer sur une autre valeur de ce même ensemble (si ce filtre peut être multivalué, nous voyons tout de même les autres valeurs également prise pour le filtre appliqué) : disons que je veux filtrer les datasets produit par le « Service des voiries » ou par le « Service des espaces verts », dès j'aurai sélectionné « Service des voiries », toutes les autres mots-clés disparaissent et ne peuvent plus être sélectionné.

Il est intéressant de constater dès qu'un filtre est appliqué, on constate que le nombre de datasets relatifs aux valeurs prises dans les autres filtres s'adapte automatiquement. De plus, les datasets peuvent être triés par date de modification, par nom, par nombre de records ou par nombre de téléchargements.

Téléchargement des données

Les datasets peuvent être téléchargés en CSV, JSON et Excel. Si les données comprennent des références géographiques, les données peuvent être exportées en GeoJSON, Shapefile et KML mais il y a une limite sur le nombre de records qu'on peut télécharger (50 000 records). Si le jeu de données contient trop de données pour un utilisateur, il peut effectuer des filtres sur celui-ci et n'exporter que le jeu de données filtré. OpenDataSoft possède pour le moment deux versions de son API dont la 2^e offre davantage de flexibilité mais elle est toujours en version beta. L'API version 1 peut filtrer (filtrage classique ou filtrage sur des données géographiques), récupérer certains champs et agréger. Les fonctionnalités sont bien présentes mais elles sont présentes sous 4 API différentes (il existe d'autres API sur le catalogue des datasets et sur les métadonnées d'un dataset ainsi que des API liées à des données géographiques) [18] :

- API de recherche d'enregistrement : récupération de données où l'idée n'est que de récupérer un subset des données, pour afficher sur une application par exemple. Voici les possibilités :
 - o Filtres ;
 - o Récupération de certains champs : l'ordre des champs en sortie ne dépend pas de l'ordre des champs indiqués dans la valeur de ce paramètre ;
 - o Récupération d'un certain nombre de records : par défaut, il est à 10 et si on veut tout récupérer, il faut mettre une très grande valeur pour ce paramètre ;
 - o Format des données en sortie : json, jsonp, geojson, geojsonp, rss, atom.
- API de téléchargements d'enregistrement : le but est de récupérer toutes les données – et selon la documentation d'OpenDataSoft, cela est plus efficace pour celui demandant les données (idée de « bulk download »). Voici les possibilités :
 - o Filtres ;
 - o Récupération de certains champs : l'ordre des champs en sortie ne dépend pas de l'ordre des champs indiqué dans la valeur de ce paramètre ;
 - o Récupération d'un certain nombre de records : par défaut, il est à 10 et si on veut tout récupérer, il faut mettre une très grande valeur pour ce paramètre ;
 - o Format des données en sortie : json, jsonp, csv, geojson, geojsonp, xls, shp, gpx, kml, ov2, geopoints, citadel, ics, dcat_ap_ch.
- API d'analyse d'enregistrement : cela est utile dans le cadre où des agrégations doivent être réalisées. Voici les possibilités :
 - o Filtres ;
 - o Choix de la/des variable(s) 'group by' ainsi de la/des variable(s) sur laquelle/lesquelles l(es) agrégations sera/seront réalisée/réalisées (ainsi que la variable d'agrégation : COUNT, AVG, SUM, MIN, MAX, STDDEV, SUMSQUARES) ;
 - o Tri possible sur la/les variable(s) group by ou la/les variable(s) d'agrégation ;
 - o Format de données en sortie : json, jsonp, csv.
- API de récupération d'un enregistrement : cette API a pour but de retrouver un record particulier au sein d'un dataset. Voici les possibilités :
 - o Format de données en sortie : json, jsonp.

Les fonctionnalités recherchées sont bien présentes mais on remarque qu'il y a un manque de clarté dans toutes ces possibilités et il est difficile de déterminer quel est le meilleur choix quand plusieurs API répondent à un même besoin. Selon la documentation présente sur OpenDataSoft, la nouvelle API (version 2) est prometteuse puisqu'elle offrira [19] :

- des nouvelles fonctionnalités de filtrages sur des données de types géo-graphiques
- des nouvelles fonctions et expressions arithmétiques (calcul de nouveaux champs sur base de champs existants)
- la possibilité d'agrégation sous un langage similaire à SQL

Il est conseillé aux développeurs d'application d'enregistrer leur application afin d'obtenir un token pour pouvoir utiliser l'API tout en étant reconnu par la plateforme [20].

Analyse de données : affichage des données brutes, tris, simples filtres et fonctionnalités avancées (graphiques, cartes, ...)

La plateforme permet de pouvoir interagir directement avec les données via une table, des tris, des filtres ainsi qu'une gamme avancée de visualisations (graphiques, images, cartes, calendrier, infographies [21]) : ces visualisations sont présentes sous la forme de différents widgets. La navigation est claire et intuitive et chacun des éléments permet de mettre les données en perspective, les traiter pour en retirer des informations pertinentes.

Upload de données : possibilités ETL avancées

Concernant la gestion de la lecture des sources de données jusqu'à la mise à jour des jeux de données sur la plateforme, OpenDataSoft offre une solution complète qui permet aux administrateurs de la plateforme d'être efficaces. Les jobs effectuant la mise à jour des jeux de données peuvent être automatisés à intervalle régulier : les intervalles peuvent être définis en mois, jour de la semaine, jour, heure. Un intervalle à la minute peut être demandé à OpenDataSoft si cela est nécessaire. Cette automatisation peut contenir plusieurs conditions, par exemple, un rafraichissement tous les mois (à minuit) avec un rafraichissement tous les lundis (à 8h du matin). Il y a également une API Push qui permet à une application d'envoyer des données en continu à la plateforme. Cela est utile quand on est fait face à des capteurs qui envoient via une API des données à très faible intervalle de temps. Dans ce cas, il est également possible d'être notifié lorsque les données ne sont plus poussées vers la plateforme à partir d'une certaine durée d'inactivité (configurable par les administrateurs de la plateforme) : cela est très utile pour détecter lorsqu'un capteur ne délivre plus de données.

La partie 'lecture de données' est puissante car la plateforme peut lire un grand nombre de formats différents : csv, tsv, URL, json, Excel, OpenDocumentSpreadsheet, Shapefile, GeoJSON, KML, KMZ ainsi que des données déjà présentes sur un portail OpenDataSoft. La limite de la plateforme est que le fichier lu doit faire au plus 240 MB mais ce fichier peut être un fichier zip. De plus, la plateforme offre une série de possibilités de transformations (appelé « processeurs ») pouvant faire du parsing de champ texte, des jointures, des transformations en données géographique à partir d'une adresse, des transpositions, des filtres, des nouveaux champs, ... : la documentation à ce sujet est exhaustive et est bien expliquée. Concernant le chargement, OpenDataSoft peut réaliser des upserts (update et insert grâce à une clé spécifiée dans le jeu de données : il n'y a pas de risque d'insérer des doublons) et des deletes.

Gestion des accès aux données et à la plateforme

La plateforme permet de gérer aussi bien les autorisations pour un utilisateur particulier que pour un groupe d'utilisateurs. Il y a 3 possibilités d'accès aux données qui peuvent être combinées et qui peuvent être facilement mises en place :

- il est possible de rendre un jeu de données visible par tout le monde ou seulement pour un ensemble d'utilisateurs et ou de groupes d'utilisateurs. Il est également possible de définir qu'un dataset soit visible par tout le monde, sauf certains utilisateurs ou groupe d'utilisateurs.
- Il est possible de rendre seulement les métadonnées d'un dataset visible que ce soit pour les utilisateurs non enregistrés ou ceux enregistrés.
- Il est possible d'afficher seulement certains records et/ou certains champs (pour tous les utilisateurs ou pour certains utilisateurs).

Création et gestion de meta-données

OpenDataSoft suit les standards en termes de métadonnées : il suit le standard DCAT.

Appliquer une charte graphique à la plateforme

L'apparence de la plateforme OpenDataSoft peut être gérée par l'administrateur de la plateforme : nom de la plateforme, url, favicon, thème utilisé par la plateforme (HTML/CSS). [22]

3.1.3.4.2.2 Should have

Faire le suivi du chargement et de l'utilisation des datasets et de l'API

OpenDataSoft a prévu une liste de jeux de données pour faire leur suivi au niveau du rafraichissement. De plus, la plateforme intègre aussi un rapport journalier qui reprend [23] :

- Les datasets en temps réels n'ayant plus reçu de données depuis une période spécifiée,
- Datasets en erreurs (qu'ils soient automatisés ou non),
- Datasets créés la veille,
- Datasets en attente d'un feedback back utilisateurs.

Les administrateurs ont aussi à leur disposition des moyens de consulter l'usage qui est fait de la plateforme. Un dashboard est prévu pour suivre l'évolution de l'utilisation de la plateforme : nombre de d'appels à l'API, nombre de visiteurs uniques, si les utilisateurs sont anonymes ou enregistrés, ceux utilisant un appareil mobile. L'administrateur peut également télécharger ces données pour qu'il réalise ses propres analyses.

Alimenter un portail à un niveau supra

La plateforme peut être moissonnée par un autre portail car les métadonnées des datasets peuvent être récupérer en appelant l'API adéquate (API de recherche de jeux de données) [24].

3.1.3.4.2.3 Could have

Interaction de la communauté autour de la donnée

Il y a plusieurs moyens au sein de la plateforme d'OpenDataSoft qui favorisent l'idée de communauté [25] :

- Les utilisateurs peuvent commenter les datasets (*peut être activé ou désactivé*).
- Les utilisateurs (enregistrés ou non) peuvent proposer les réutilisations qui ont été faites sur base d'un jeu de données (*peut être activé ou désactivé*) et les administrateurs approuvent ces réutilisations.
- Les utilisateurs peuvent directement cliquer sur les widgets liés aux réseaux sociaux (Twitter, Facebook, Google+, LinkedIn et mail) pour qu'ils puissent partager avec leur réseau un lien pointant vers le dataset (*peut être activé ou désactivé*).
- Les utilisateurs peuvent poser leur question via un formulaire.
- Les utilisateurs enregistrés peuvent être informés de modifications importantes concernant un dataset particulier ainsi le producteur de données peut prévenir les utilisateurs si des changements majeurs sont prévus.
- Une version beta pour la « contribution des utilisateurs à un jeu de donnée » est possible – non disponible sur la plateforme par défaut – et elle permet aux utilisateurs de donner un feedback sur les valeurs prises par un record particulier ou pour proposer un nouveau record mais c'est bien le producteur de données qui doit s'assurer de la pertinence des feedbacks et de faire les modifications nécessaires dans les données.

Enregistrer/exporter les visualisations réalisées par l'utilisateur

Les graphiques, les tableaux, les cartes, les infographies affichés à l'écran peuvent être facilement réintégrés dans d'autres sites web en faisant simplement un copier-coller du code HTML généré.

Support technique sur la plateforme pour les utilisateurs (tutoriels, ...)

La documentation d'OpenDataSoft est bien structurée et très didactique, aussi bien pour les développeurs que pour ceux étant moins techniques.

Possibilité que le portail puisse extraire les meta-données sur un autre portail

Le portail peut également aller puiser les meta-données depuis d'autres portails : cela permet de référencer les données d'un autre portail sur notre portail. Voici les moissonneurs les plus utiles : moissonneur ArcGIS, moissonneur CKAN, moissonneur Junar, moissonneur Socrata. Ces moissonneurs peuvent également mettre à jour les meta-données à intervalle régulier : la seule limitation est qu'ils ne peuvent être exécutés au maximum une fois par jour.

3.1.3.4.3 Tarification

La tarification chez OpenDataSoft dépend des éléments suivants

- du nombre de datasets,
- du volume de l'ensemble des datasets,
- du nombre total de lignes qu'un dataset peut atteindre,
- du volume maximum du contenu non structuré (images, pièces jointes aux jeux de données, ...),
- du nombre d'appels à l'API par mois,

- Du volume maximum du données mises à jour par jour.

Toutes les options de la plateforme sont disponibles quel que soit la licence choisie : OpenDataSoft ne vend pas sa plateforme sous forme de modules.

La licence de base, proposée à 11 000€ par an, convient si la plateforme contient moins de 75 datasets et que la taille de ceux-ci n'excède pas 5 GB. Si davantage est requis, la licence suivante s'élève à 22 000€ par an si la plateforme contient moins de 150 datasets pour un volume total de 10 GB.

3.1.4 Comparaison des plateformes et recommandations

3.1.4.1 Discussion des véritables différences entre les plateformes

Les 2 plateformes seront quantitativement notées : on attribuera un score de 0 à 5 pour chaque besoin précédemment analysé. Un poids sera également attribué pour chaque besoin. Cela nous aidera à avoir une meilleure vue sur l'ensemble de ces critères ainsi qu'à pointer les différences notables entre Socrata et OpenDataSoft. Seulement les différences d'au moins 1, en valeur absolue, seront commentées.

		<i>poids</i>	Socrata	OpenDataSoft	<i>Différence</i>
MUST HAVE	Présent sur le Web	20	5	5	0.00
	Recherche	20	4.5	4.5	0.00
	Download données	60	4.5	3.5	-1.00
	Analyse de données	40	3	4	1.00
	ETL	30	2	4.5	2.50
	Gestion accès données et utilisateurs	20	2.5	4	1.50
	Meta-données	25	4	4	0.00
	Customiser plateforme	10	3.5	3.5	0.00
	Sous-total		3.67	4.04	
SHOULD HAVE	Suivi upload et utilisation	30	4	4	0.00
	Alimenter un portail supra	20	3.5	3.5	0.00
	Sous-total		3.80	3.80	
COULD HAVE	Communauté	20	3.5	4	0.50
	Exportation/réutilisations visualisations	15	4	4	0.00
	Support pour utilisateurs/administrateurs	30	3	4	1.00
	Puier dans un portail	10	4	4	0.00
	Sous-total		3.47	4.00	
TOTAL			3.64	4.00	

Au niveau du « Download données », Socrata est meilleur puisqu'il offre systématiquement une gamme plus large de format de téléchargement de données. Néanmoins, OpenDataSoft offre également les données sous format non propriétaire : il répond donc au besoin. Socrata possède une API simple d'utilisation et très robuste : il n'y a qu'une seule manière de l'appeler alors que chez OpenDataSoft, il y a plusieurs manières de l'appeler en fonction de l'objectif poursuivi. Les fonctionnalités qu'on attend de l'API sont présentes aussi chez OpenDataSoft. Pour conclure sur ce besoin, OpenDataSoft répond à ce besoin mais il y répond avec une aisance moindre que ce que fournit Socrata : voilà pourquoi Socrata a un score de 4,5 alors qu'OpenDataSoft obtient un score de 3,5.

L'analyse de données est à l'avantage d'OpenDataSoft : l'interface pour jouer avec les données dans un tableau est facile d'utilisation et les graphiques sont facilement malléables. Il est possible d'afficher des graphiques par défaut pour que l'utilisateur puisse déjà bénéficier d'une valorisation des données. Au sein de la plateforme Socrata, il est également possible de jouer avec les données dans une table ou de réaliser les graphiques mais l'interface a un côté plus technique, moins grand public. En rappelant que les objectifs du BEP sont la transparence ainsi que la sensibilisation/éducation du citoyens, OpenDataSoft répond donc mieux à ces objectifs.

La différence notable entre les 2 plateformes se situe sur l'aspect « ETL » : un aspect primordial pour les futurs administrateurs de la plateforme. OpenDataSoft offre une multitude d'outils permettant d'automatiser la lecture de données afin d'alimenter la plateforme : il contient toute une série de fonctionnalités permettant de maintenir à jour efficacement sa plateforme via un ETL puissant. Comme il l'a été décrit plus haut, Socrata a besoin d'autres outils pour automatiser le rafraichissement de données : ces autres outils nécessitent une gestion interne pour leur maintenance. Socrata amène une complexité et une maintenance interne qui sera synonyme de coût supplémentaire. OpenDataSoft est donc clairement un meilleur parti pour offrir une autonomie et une efficacité aux administrateurs de la plateforme.

OpenDataSoft excelle dans la gestion des accès aux données : il peut rendre un jeu de données privé ou public et il peut même spécifier les utilisateurs ayant accès seulement à certaines lignes et/ou à certains champs présents dans le jeu de données. Socrata peut uniquement rendre un dataset public ou privé. OpenDataSoft obtient donc la note de 4 tandis que Socrata obtient la note de 2,5.

La documentation pour les utilisateurs ou les administrateurs de la plateforme est davantage détaillée et vulgarisée pour OpenDataSoft que pour Socrata. OpenDataSoft obtient un score de 4 tandis que Socrata obtient un score de 3.

3.1.4.2 Recommandation

Socrata est un outil dont le périmètre d'action est plus limité mais les fonctionnalités implémentées sont hautement avancées. Il s'inscrirait davantage dans des organisations dont la maturité avec les données est plus grande. Socrata a plutôt l'air de se tourner vers un public plus technique où la réutilisation de données est majoritairement réalisée via l'appel à son API : Socrata prône l'utilisation de son API pour favoriser le développement d'applications.

OpenDataSoft offre une navigation et une expérience d'utilisation à portée d'un public large et non-expert. OpenDataSoft est axé sur la réutilisation des données au sein de la plateforme. L'expérience de navigation est fortement empreinte de transparence et d'accompagnement envers l'utilisateur.

Sur base de l'analyse et des objectifs que le BEP poursuit (transparence, approche Smart City, engagement des citoyens et autonomie), la plateforme OpenDataSoft semble la plus adéquate pour le BEP.

3.2 Prochaines étapes

Le BEP pourra utiliser ce travail pour réaliser un cahier des charges ainsi qu'un appel d'offres. Les membres avec lesquels j'ai travaillé sont en train de répandre le mot au sein du management mais également au sein des départements pour parler du phénomène de l'Open Data.

Plusieurs points sont à considérer encore dans une approche Open Data : les licences, la réutilisation des données et la protection des données à caractère personnel. Ces points sont abordés de manière générique et introductive pour donner aux parties prenantes matière à débiter.

3.2.1 Licences

Les licences permettent de préciser les conditions, les droits et éventuellement les obligations de l'utilisateur de données. En effet, les bases de données « dont l'obtention, la vérification ou la présentation du contenu atteste un investissement qualitativement ou quantitativement substantiel » (Art. XI.306. [26]) bénéficient d'un droit sui generis : le droit sui generis protège la base de données dans son ensemble (et non les éléments pris individuellement) [27]. Le but est de protéger l'investissement que le producteur a réalisé dans sa base de données. Ce n'est pas parce que les données sont mises sur un site web que le producteur de la base de données renonce à ses droits. Plusieurs cas de figures se présentent :

- Si l'extraction de la base de données est substantielle, le producteur de la base de données ne peut pas faire valoir ses droits si le but de l'extraction est strictement privé, à des fins d'enseignement ou de recherche scientifique à but non lucratif ou à des fins de sécurités publiques, procédures administratives ou juridictionnelles.
- Si l'extraction est non substantielle, le producteur de la base de données ne peut pas faire valoir ses droits pour autant que l'extraction n'ait pas de caractère répété.

Une licence ouverte protège les utilisateurs sur les réutilisations qu'ils peuvent faire des données. Néanmoins, 2 obligations sont souvent maintenues : citer leur source (*attribution*) et publier les données qu'ils en ont dérivées (*share-alike*) [28]. Il y a donc 3 licences ouvertes standards [28] :

- Licence du domaine public : il n'y a aucune obligation à remplir par le réutilisateur des données.
- Licence de « type attribution » : l'utilisateur doit mentionner la source des données.
- Licence de « type attribution et share-alike » : l'utilisateur doit mentionner la source des données et partager les données qu'il en a dérivée sous la même licence que l'utilisateur en a bénéficié.

Open Data Institute conseille d'utiliser des licences standards disponible sur le web car cela représente moins de travail, on est sûr que le langage juridique est adéquat et il est plus facile pour les utilisateurs d'identifier ce qu'ils peuvent faire avec les données [28]. Ces licences ne doivent pas limiter les conditions de réutilisations des documents publics (Art. 4, décret du Parlement wallon relatif à la réutilisation des informations du secteur public et visant à l'établissement d'une politique de données ouvertes (« Open Data ») pour les matières visées par l'article 138 de la Constitution).

3.2.2 Cadres réglementaires

3.2.2.1 Réutilisation des données du secteur public

Le but de cette section est de rappeler le cadre réglementaire qui entoure l'accès et la réutilisation des données publiques.

Ces réglementations qui encadrent la réutilisation des données publiques ont été initialement lancées en 2003 par la directive 2003/98/CE venant du Parlement Européen. Cette directive déterminait les conditions auxquelles les pays pouvaient mettre les données publiques à disposition. Cette directive n'avait pas pour but d'obliger les pays à ouvrir les données publiques [29]. Il s'est avéré que l'effet économique escompté n'avait pas été atteint et ils ont réalisé qu'il pouvait aller plus loin, par une nouvelle directive (2013/37/UE).

« Les principales modifications apportées par la directive 2013/37 portent sur l'obligation pour les organismes du secteur public d'autoriser la réutilisation des documents qui sont accessibles au public, sur le champ d'application de la directive, sur la mise à disposition des métadonnées, sur le calcul des redevances par les administrations, ainsi que sur des modalités techniques et pratiques destinées à faciliter la réutilisation. » [29] Ces données peuvent être à disposition du public pour autant que cela ne soit pas en violation d'autres dispositions européennes ou nationales (vie privée, sécurité nationale, droits de propriété intellectuelle).

La directive 2013/37 fut transposée au niveau fédéral le 4 mai 2016. Il y a donc clairement un nouveau paradigme : l'ouverture des données est dorénavant la règle [30]. Dans le plan stratégique défini par Alexander de Croo, l'objectif est d'arriver à l'« open by default » [30].

La région wallonne a voté un décret le 12 juillet 2017 qui a pour but de transposer également la directive 2013/37 tout en mettant en place une stratégie Open Data harmonisée pour le territoire [31]. Ce décret rentre en vigueur le 1^{er} septembre 2017 (Art. 13). Voici les points majeurs de ce décret :

- Art. 6 §1^{er}, alinéa 1 : « Les documents administratifs mis à disposition par les organismes publics sont répertoriés et publiés sur le portail commun dédié à la réutilisation des informations du secteur public géré par l'Agence du Numérique. »
- Art. 11 §1^{er}, alinéa 1 : « Le document administratif disponible en vue d'une réutilisation, les conditions éventuelles dont les licences types, ainsi que le montant des rétributions éventuelles en ce compris la base de calcul, sont répertoriés et publiés, sur le portail commun. »

On voit donc clairement que les cadres réglementaires ainsi que la stratégie au niveau fédéral et régional sont finalisés et sur le point d'être en application.

3.2.2.2 Protection des données à caractère personnel

L'ouverture des données publiques doit bien évidemment se faire dans le respect de la loi relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel (loi du 8 décembre 1992) : « Des données à caractère personnel sont seulement réutilisables dans la mesure où cette réutilisation n'est pas incompatible avec la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel et ses arrêtés d'exécution, ainsi qu'avec la loi du 15 janvier 1990 portant création et organisation d'une Banque carrefour de la sécurité sociale. » (Art. 3 §3, loi relatif à la réutilisation des informations du secteur public du 4 mai 2016).

La délicate mission d'évaluation de risques lorsqu'une autorité veut ouvrir des données à caractère personnel incombe au comité sectoriel PSI (Public Sector Information) de la Commission de protection de la vie privée (Art. 3 §3, alinéa 2, loi relatif à la réutilisation des informations du secteur public du 4 mai 2016). Ce comité est là pour conseiller, guider les autorités de façon à minimiser le risque de réidentification.

A partir du 25 mai 2018, les traitements des données à caractère personnel devront suivre le Règlement Général de la Protection des Données (RGPD). Toute autorité ou service « qui, seul ou conjointement avec d'autres, détermine les finalités et les moyens du traitement [de données à caractère personnel] » est responsable de traitement (Art. 4 RGPD). Le responsable de traitement doit être en mesure de satisfaire aux exigences légales imposées par le RGPD. Voici quelques-unes des exigences issues du RGPD :

- Tenue d'un registre interne des traitements de données à caractère personnel (Art. 30 RGPD)
- Coopération avec l'autorité de contrôle (Art. 31 RGPD)
- Traitement licite (Art. 6 RGPD), loyal et transparent (Art. 12 et 13 RGPD)
- Données pertinentes et limitées au regard des finalités poursuivies (Art. 5 RGPD)
- Donner une vue à la personne concernée sur l'utilisation qu'il est fait de ses données personnelles (Art. 15 RGPD)
- Exactitude des données (Art. 16 RGPD)
- Droit à l'oubli (Art. 17 RGPD)
- Droit à la portabilité des données (Art. 20 RGPD)

Sauf si une des 6 conditions autorisant le traitement de données à caractère personnel est remplie (Art. 6 RGPD), ouvrir des données à caractère personnel est considéré comme illicite [32]. De plus, le RGPD possède 2 principes difficilement conciliables avec la logique venant de l'Open Data [32] :

- Les finalités poursuivies doivent être définies et spécifiques : avec l'Open Data, les finalités poursuivies sont déterminées par les réutilisateurs et elles peuvent être très diverses.
- Si les données traversent les frontières, alors celui qui utilise les données doit montrer un niveau de protection adéquat : il est matériellement impossible de contrôler tous les réutilisateurs des données en matière d'Open Data.

Une solution pour publier des données à caractère personnel est de ne plus tomber sous le sceau du champ d'application du RGPD, c'est-à-dire anonymiser les données de sorte à ce que cela ne soit plus des données à caractère personnel [32]. Anonymiser les données est également un défi car il faut non seulement pouvoir créer un jeu de données qu'on ne peut plus lier à un individu mais il faut pouvoir également ne pas être en mesure de lier ces données avec des données tierces (que d'autres possèdent ou qui sont également disponibles en Open Data), tout en gardant à l'esprit que ces données tierces ne font que s'accroître avec le temps et donc que les possibilités de réidentification ne sont que plus grandes [32].

D'autres techniques telles que la pseudonimisation existent mais le but de cette méthode est de pouvoir lier les données entre datasets tout en ne sachant pas identifier directement quel est l'individu concerné. Ces liens entre données augmentent la quantité d'information sur l'individu et augmente donc la probabilité de l'identifier [32].

Les données anonymisées ont toujours le risque d'être désanonymisées avec suffisamment de données tierces : « le risque de réidentification ne peut être réduit à zéro » [32]. La logique qui va donc prévaloir lorsqu'on mesure le risque de réidentification est d'avoir un coût de réidentification qui est plus important que les bénéfices d'avoir des données non-anonymisées [32].

4 Conclusion

Ce travail est parti d'une collaboration avec le Bureau Economique de la Province de Namur. Le BEP avait déjà réfléchi à voir ce qu'il pouvait faire en Open Data mais il n'était pas encore rentré dans un travail de fond à ce sujet. Le BEP avait déjà une stratégie au niveau des Smart City et le but de ce travail a été de mettre noir sur blanc une stratégie potentielle d'Open Data en lien avec sa stratégie Smart City ainsi qu'adaptée au contexte économique, social et politique actuel. Cette stratégie potentielle trouva 2 sources majeures : la littérature sur le sujet ainsi que les discussions et réflexions menées en groupe pour comprendre comment le BEP fonctionne. La façon de construire la stratégie ainsi que la stratégie en tant que telle reflète l'identité du BEP : autonomie des parties ainsi que transparence.

Dans la deuxième partie du projet, la stratégie fut résumée de manière succincte. En effet, il fallait donc avoir une vue sur les objectifs à atteindre avec la plateforme Open Data. Cette stratégie s'est constituée de manière plus concrète et plus tangible lorsque l'identification des besoins a été réalisée. Cette stratégie a été également notre grille de lecture nous permettant de prioriser les besoins en 3 groupes distincts. L'identification et la priorisation des besoins se sont réalisés grâce à plusieurs itérations.

Après que chaque besoin ait été décrit, Socrata et OpenDataSoft furent analysées. Cet analyse fut réalisée, dans la mesure du possible, en situation concrète en utilisant des plateformes Open Data disponibles sur le Web. Après une description détaillée de chacune des plateformes, elles furent évaluées et une discussion des différences suivit. Il s'est avéré que la plateforme OpenDataSoft correspond davantage aux objectifs de notre stratégie potentielle ainsi qu'à la stratégie IT définie par le département IT du BEP. OpenDataSoft peut offrir la transparence, l'engagement des citoyens, le professionnalisme que requiert une approche Smart City ainsi que l'autonomie des personnes administrant la plateforme.

Nous n'oublions pas que la plateforme n'est aussi « qu'un outil » et qu'il est dès lors nécessaire de réfléchir aux données à mettre à disposition ainsi qu'aux moyens à mettre en œuvre pour alimenter la plateforme Open Data. Le BEP travaille déjà sur sa « culture data » puisqu'elle est en possession d'un outil de Business Intelligence : le BEP est une entreprise publique moderne qui est orientée solutions et elle se base sur des faits mesurables et vérifiables. Comme il a déjà été mentionné, le BEP doit encore convaincre ses membres ainsi qu'augmenter la conscientisation de l'Open Data dans les communes et chez les citoyens. Des initiatives sont encore nécessaires pour faire grandir l'Open Data en province de Namur. Le but ultime est de montrer l'utilité et de faire bénéficier aux citoyens des impacts générés à partir de l'Open Data.

5 Bibliographie

- [1] Bureau Economique de la Province de Namur, «Le BEP fête ses 50 ans,» Namur, 2013.
- [2] Wallex, 11 décembre 1986. [En ligne]. Available: <https://wallex.wallonie.be/index.php?doc=5902&rev=5181-2821>. [Accès le 3 août 2017].
- [3] Bureau Économique de la Province de Namur, «Culture d'entreprise,» [En ligne]. Available: <http://www.bep.be/>. [Accès le 19 Février 2017].
- [4] Bureau Économique de la Province de Namur, «Plan Stratégique 2017,» 2017.
- [5] Bureau Économique de la Province de Namur, *Open Data - Pourquoi et comment partager les données de ma commune?*, Marche-en-Famenne, 2016.
- [6] W. Carrara, W. Engbers, M. Nieuwenhuis et E. van Steenberg, «Analytical Report 4: Open Data in Cities,» 2016.
- [7] Open Data Institute, «Guides: What makes data open?,» Open Data Institute, [En ligne]. Available: <http://www.theodi.org/guides/what-open-data>. [Accès le 5 Mars 2017].
- [8] W. Carrara, M. Nieuwenhuis et H. Vollers, «Open Data Maturity,» European Commission, 2016.
- [9] W. Carrara, W. S. Chan, S. Fischer et E. van Steenberg, «Creating Value Through Open Data,» European Union, 2015.
- [10] Ville de Paris, «Paris renforce sa démarche pour capter et exploiter la donnée,» 4 avril 2016. [En ligne]. Available: http://www.paris.fr/services-et-infos-pratiques/innovation-et-recherche/ville-intelligente-et-durable/paris-renforce-sa-demarche-pour-capter-et-exploiter-la-donnee-3191#les-autres-experimentations-en-cours_5. [Accès le 2 avril 2017].
- [11] P. Weisberg, «DataSF Research Review: Open Data Evaluation Framework,» DataSF.org, 2014.
- [12] City of New York, «Open Data Policy and Technical Standards Manual,» 2016.
- [13] «8 principes,» 7 Décembre 2007. [En ligne]. Available: https://public.resource.org/8_principles.html. [Accès le 16 Mars 2017].
- [14] W3C, «Data Catalog Vocabulary (DCAT),» 16 Janvier 2014. [En ligne]. Available: <https://www.w3.org/TR/vocab-dcat/>. [Accès le 22 Avril 2017].
- [15] Socrata, «Catalog Search,» [En ligne]. Available: <https://support.socrata.com/hc/en-us/articles/225465147-Catalog-Search-FAQ>. [Accès le 21 Mai 2017].
- [16] Socrata, «<https://dev.socrata.com/docs/queries/>,» [En ligne]. [Accès le 21 Mai 2017].

- [17] Socrata, «Publica Release Notes for 5/17/2017: Data.JSON Metadata Federation,» [En ligne]. Available: <https://support.socrata.com/hc/en-us/articles/115007458088>. [Accès le 22 Mai 2017].
- [18] OpenDataSoft, «Documentation - API des enregistrements des jeux de données,» [En ligne]. Available: https://docs.opendatasoft.com/fr/using_api/v1/records_api.html. [Accès le 25 Mai 2017].
- [19] OpenDataSoft, «Documentation - Utiliser l'API,» [En ligne]. Available: https://docs.opendatasoft.com/fr/using_api/introduction.html. [Accès le 25 Mai 2017].
- [20] OpenDataSoft, «Documentation - Utiliser l'autorisation OAuth2,» [En ligne]. Available: https://docs.opendatasoft.com/fr/using_api/oauth2.html. [Accès le 25 Mai 2017].
- [21] Hauts-de-Seine, «Résultats des élections présidentielles par commune,» [En ligne]. Available: <https://opendata.hauts-de-seine.fr/explore/dataset/resultats-des-elections-par-commune/infographie/>. [Accès le 25 Mai 2017].
- [22] OpenDataSoft, «Documentation - Personnaliser l'apparence du portail,» [En ligne]. Available: http://docs.opendatasoft.com/fr/customizing_look_and_feel/index.html. [Accès le 26 Mai 2017].
- [23] OpenDataSoft, «Documentation - Rapport journalier,» [En ligne]. Available: http://docs.opendatasoft.com/fr/notifications/daily_report.html#daily-report. [Accès le 26 Mai 2017].
- [24] OpenDataSoft, «Documentation - API du catalogue de jeux de données,» [En ligne].
- [25] OpenDataSoft, «Documentation - Faire participer les utilisateurs,» [En ligne]. Available: http://docs.opendatasoft.com/fr/configuring_portal/engaging.html. [Accès le 26 Mai 2017].
- [26] SPF Economie, Code de Droit Economique, Livre XI.
- [27] SPF Economie, «La protection des bases de données par la propriété intellectuelle,» [En ligne]. Available: http://economie.fgov.be/fr/entreprises/propriete_intellectuelle/droit_bases_de_donnees/protection_bases_de_donnees. [Accès le 5 août 2017].
- [28] Open Data Institute, «Publisher's Guide to Open Data Licensing,» [En ligne]. Available: <https://theodi.org/guides/publishers-guide-open-data-licensing>. [Accès le 5 août 2017].
- [29] C. Ker, «Réutilisation des informations du secteur public : la transposition de la directive 2013/37/UE,» [En ligne]. Available: <http://www.rdti.be/Doc-Jur/doc62-2.html>. [Accès le 6 août 2017].
- [30] data.gov.be, «La Chambre votera demain une ambitieuse loi open data,» 13 avril 2016. [En ligne]. Available: <http://data.gov.be/fr/news/la-chambre-votera-demain-une-ambitieuse-loi-open-data>. [Accès le 6 août 2017].

- [31] Union des Villes et des Communes de Wallonie, «Opendata – vote par le Parlement wallon du décret conjoint relatif à la réutilisation des informations du secteur public et visant à l'établissement d'une politique de données ouvertes,» Union des Villes et des Communes de Wallonie, 24 juillet 2017. [En ligne]. Available: <http://www.uvcw.be/actualites/33,210,39,39,7152.htm>. [Accès le 7 août 2017].
- [32] E. Simperl, K. O'Hara et R. Gomer, «Analytical Report n3: Open Data and Privacy,» European Data Portal.
- [33] I. Sussha, A. Zuiderwijk, Y. Charalabidis, P. Parycek et M. Janssen, «Critical Factors for Open Data Publication and Use: A Comparison of City-level, Regional, and Transnational Cases,» *Journal of Democracy*, 2015.